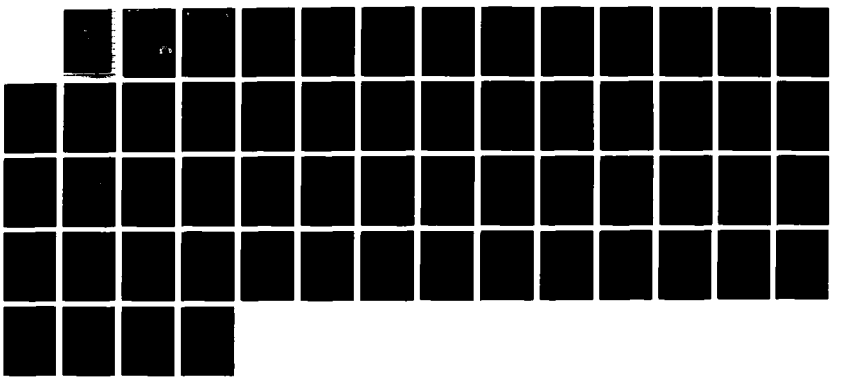
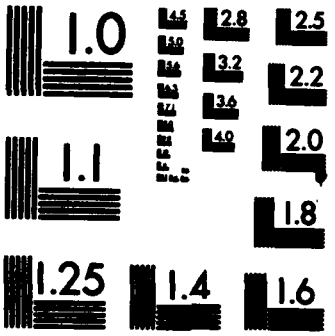
AD-A181 500 NAVAL C3 (COMMAND-AND-CONTROL-AND-COMMUNICATIONS) 1/1
DISTRIBUTED TACTICAL DE..(U) MASSACHUSETTS INST OF TECH
CAMBRIDGE LAB FOR INFORMATION AND D.. M ATHANS ET AL.
UNCLASSIFIED 15 APR 87 LIDS-IR-1669 N00014-84-K-0519 F/G 25/5 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

AD-A181 500

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NAVAL C^3 DISTRIBUTED TACTICAL DECISION MAKING		5. TYPE OF REPORT & PERIOD COVERED 1 April 1986 - 31 March 1987 6. PERFORMING ORG. REPORT NUMBER LIDS-IR-1669
7. AUTHOR(s) Michael Athans Alexander H. Levis		8. CONTRACT OR GRANT NUMBER(s) N00014-84-K-0519
9. PERFORMING ORGANIZATION NAME AND ADDRESS Laboratory for Information and Decision Systems Massachusetts Institute of Technology Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 649-003
11. CONTROLLING OFFICE NAME AND ADDRESS ENGINEERING PSYCHOLOGY PROGRAMS (CODE 442EP) Office of Naval Research Arlington, Virginia 22217-5000		12. REPORT DATE April 15, 1987 13. NUMBER OF PAGES i+34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Mr J.D. Simpson Code 111SP Mathematical Sciences Division Office of Naval Research Arlington, Virginia 22217-5000		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTED JUN 18 1987 S E D

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

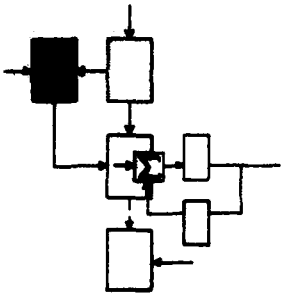
Command and Control, Distributed Decisionmaking, Organization Theory,

Petri Nets

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Progress on eight research problems addressing distributed tactical decisionmaking is described.

DD FORM 1473



LABORATORY FOR
INFORMATION AND DECISION SYSTEMS

Massachusetts Institute of Technology
Cambridge, Mass., 02139, U.S.A.

OSP NUMBER 95442

LIDS-IR-1669

ANNUAL TECHNICAL REPORT

for the period

1 April 1986 to 31 March 1987

for

NAVAL C^3 DISTRIBUTED TACTICAL DECISION MAKING

Contract Number N00014-84-K-0579
Work Unit NR 649-003

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

Submitted to:

Mr. J. R. Simpson, Code 111SP
Dr. L. Broemeling, Code 111SP
Dr. W. R. Vaughan, Jr.

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Submitted by:

Michael Athans
Alexander H. Levis

April 15, 1987

SERVOMECHANISMS LABORATORY
1939-1959
ELECTRONIC SYSTEMS LABORATORY
1959-1978

87 6 16 002

NAVAL C^3 DISTRIBUTED TACTICAL DECISIONMAKING

1. PROJECT OBJECTIVES

The objective of this research is to address analytical and computational issues that arise in the modeling, analysis and design of distributed tactical decisionmaking. The research plan has been organized into two highly interrelated research areas:

(a) Distributed Tactical Decision Processes; and
(b) Distributed Organization Design.

The focus of the first area is the development of methodologies, models, theories and algorithms directed toward the derivation of superior tactical decision, coordination, and communication strategies of distributed agents in fixed organizational structures. The framework for this research is normative.

The focus of the second area is the development of a quantitative methodology for the evaluation and comparison of alternative organizational structures or architectures. The organizations considered consist of human decisionmakers with bounded rationality who are supported by C^3_{μ} systems. The organizations function in a hostile environment where the tempo of operations is fast; consequently, the organizations must be able to respond to events in a timely manner. The framework for this research is descriptive. Keywords: Petri nets; organization theory;

2. STATEMENT OF WORK

The research program has been organized into seven technical tasks -- four that address primarily the theme of distributed tactical decision processes and three that address the design of distributed organizations. An eighth task addresses the integration of the results. They are:

2.1 Real Time Situation Assessment: Static hypothesis testing, the effect of human constraints and the impact of asynchronous processing on situation assessment tasks will be explored.

2.2 Real Time Resource Allocation: Specific research topics include the use of algebraic structures for distributed decision problems, aggregate solution techniques and coordination.

2.3 Impact of Informational Discrepancy: The effect on distributed decisionmaking of different tactical information being available to different decisionmakers will be explored. The development of an agent model, the modeling of disagreement, and the formulation of coordination strategies to minimize disagreement are specific research issues within this task.

2.4 Constrained Distributed Problem Solving: The agent model will be extended to reflect human decisionmaking limitations such as specialization, limited decision authority, and limited local computational resources. Goal decomposition models will be introduced to derive local agent optimization criteria. This research will be focused on the formulation of optimization problems and their solution.

2.5 Evaluation of Alternative Organizational Architectures: This task will address analytical and computational issues that arise in the construction of the generalized performance-workload locus. This locus is used to describe the performance characteristics of a decisionmaking organization and the workload of individual decisionmakers.

2.6 Asynchronous Protocols: The use of asynchronous protocols in improving the timeliness of the organization's response is the main objective of this task. The tradeoff between timeliness and other performance measures will be investigated.

2.7 Information Support Structures: In this task, the effect of the C^3 system on organizational performance and on the decisionmaker's workload will be studied.

2.8 Integration of Results: A final, eighth task, is included in which the various analytical and computational results will be interpreted in the context of organizational bounded rationality.

3. STATUS REPORT

In the context of the first seven tasks outlined in Section 2, a number of specific research problems have been formulated and are being addressed by graduate research assistants under the supervision of project faculty and staff. Research problems which were completed prior to or were not active during this last year have not been included in the report.

3.1 DISTRIBUTED TEAM HYPOTHESIS TESTING WITH SELECTIVE COMMUNICATIONS

Background: In Command-and-Control-and-Communication(C3) systems multiple hypothesis-testing problems abound in the surveillance area. Targets must be detected and their attributes must be established; this involves target discrimination and identification. Some target attributes, such as location, are best observed by sensors such as radar. More uncertain target locations are obtained by passive sensors, such as sonar or IR sensors. However, target identity information requires other types of sensors (such as ESM receivers, IR signature analysis, human intelligence etc). As a consequence, in order to locate accurate and identify a specific target out of a possibly large potential population (including false targets) one must design a detection and discrimination system which involves the fusing of information from several different sensors generating possibly specialized information about the target. These sensors may be collocated on a platform (say a ship in a Naval battle group) and be physically dispersed as well (ESM receivers exist in every ship, aircraft, and submarine). The communication of information among this diverse sensor family may be difficult (because of EMCON restrictions) and is vulnerable to enemy countermeasure actions (physical destruction and jamming). It is this class of problems that motivates our research agenda.

Research Goals: We are conducting research on distributed multiple hypothesis testing using several decision-makers, and teams of decision-makers, with distinct private information and limited communications. The goal of this research is to *unify* our previous research in situation assessment, distributed hypothesis testing, and impact of informational discrepancy; and to *extend* the methodology, mathematical theory and computational algorithms so that we can synthesize and study more complex organizational structures. The solution of this class of basic

research problems will have impact in structuring the distributed architectures necessary for the detection, discrimination, identification and classification of attributes of several targets (or events) by a collection of distinct sensors (or dispersed human observers).

The objective of the distributed organization will be the resolution of several possible hypotheses based on many uncertain measurements. Each hypothesis will be characterized by several attributes. Each attribute will have a different *degree of observability* to different decision makers or teams of decision makers; in this manner, we shall model different specialization expertise associated with the detection and resolution of different phenomena. Since each hypothesis will have several attributes, it follows that in order to reliably confirm or reject a particular hypothesis, two or more decision-makers (or two or more *teams* of decision-makers) will have to pool and fuse their knowledge.

Extensive and unecessary communication among the decision-makers will be discouraged by explicitly assigning costs to certain types of communication. In this manner, we shall seek to understand and isolate which communications are truly vital in the organizational performance; the very problem formulation will discourage communications whose impact upon performance is minimal. Quantitative tradeoffs will be sought.

Another feature which will be incorporated relates to the vulnerability of the distributed decision process to enemy countermeasures. Thus, in our distributed decision models we shall assume that there is a finite probability that the actions (decisions and/or conclusions) of any one particular decision-maker will be distorted or destroyed due to enemy action. As a consequence, the organization of the decision teams, the protocols, and the decision rules must explicitly take into account the vulnerability issue. As a minimum, a certain level of decision-making redundancy must exist in the distributed organization; the coordination strategies and the protocols that isolate "damaged" decision-makers will be developed. We shall seek to determine, in a quantitative setting, the minimum required level of decision-maker redundancy as a function of the degree of vulnerability to enemy countermeasures (such as jamming).

We stress that we shall strive to design distributed organizational architectures in which teams of teams of decision-makers interact. For example, a team may consist of a *primary* decision-maker

together with a *consulting* decision-maker -- the paradigm used by Papastavrou and Athans.

The methodology that we plan to employ will be mathematical in nature. To the extent possible we shall formulate the problems as mathematical optimization problems. Thus, we seek *normative solution concepts*. To the extent that human *bounded rationality* constraints are available, these will be incorporated in the mathematical problem formulation. In this case, the nature of the results will correspond to what is commonly refered to as *normative/descriptive* solutions. Therefore, we visualize a dual benefit of our basic research results. From a purely mathematical point of view, the research will yield nontrivial advances to the distributed hypothesis-testing problem; an extraordinary difficult problem from a mathematical point of view. From a psychological perspective, we hope that the normative results will suggest *counterintuitive* behavioral patterns of -- even perfectly rational -- decision- makers operating in a distributed tactical decision-making environment; these will set the stage for designing empirical studies and experiments and point to key variables that should be observed, recorded and analyzed by cognitive scientists. From a military C3 viewpoint, the results will be useful in structuring distributed architectures for the surveillance function.

Progress to Date: Research was initiated in September 1987. At present we are in the modeling and problem formulation phase. The challenge is to pose the problem in such a way so that its generic richness is preserved, yet having a chance for mathematical solutions which will provide insight.

We have developed a simple model for capturing the effects of countermeasures. Suppose that we have a decision-maker that makes a binary decision, i.e. YES, I believe that I see a target vs NO, I do not believe that a target is there. We can have a small but finite probability that when the decision-maker meant to say YES the other team members hear NO, and vice versa. The degree of the countermeasures intensity can be quantified by the numerical value of the assigned probability. This way of modeling the impact of enemy countermeasures does not complicate the mathematics very much in the distributed hypothesis-testing algorithms.

Progress during the past quarter: In the past quarter we focused our attention to the problem of *ternary* hypothesis testing by a team of two cooperating decision makers; communication

between the two decision-makers is costly and consists of a finite alphabet. The problem is to distinguish among three different hypotheses. Each decision-maker obtains an uncertain measurement of the true hypothesis. The so-called *primary* decision-maker has the option of making the final team decision or consulting, at a cost, the *consulting* decision-maker. The consulting decision-maker is constrained to provide information using a ternary alphabet. The team objective is to minimize the probability of error together with the communications cost (if any).

This seemingly simple distributed decision problem turns out to have an extraordinarily complex structure. We have been able to characterize the nature of the optimal solution; however, we have not obtained as yet the formal mathematical solution and associated algorithms. Nonetheless, it is possible to obtain a significant insight into the complexity of multiple hypothesis-testing problems. Also, we have made progress in pinpointing what we mean by calling a decision-maker an "expert" in some hypotheses and a "novice" in others. These are critical issues when we examine more complex decision organizations with several members.

Many more mathematical models and tentative approaches remain to be developed. This research will most probably form the core of the Ph.D. research of J. Papastavrou under the supervision of Professor Athans.

Documentation: None as yet. A presentation is in preparation for delivery at the C^3 Symposium in June 1987.

3.2 DISTRIBUTED HYPOTHESIS TESTING WITH MANY AGENTS

Background: The goal of this research project is to develop a better understanding of the nature of the optimal messages to be transmitted to a central command station (or fusion center) by a set of agents who receive different information on their environment. In particular, we are interested in solutions of this problem which are tractable from the computational point of view. Progress in this direction has been made by studying the case of a large number of agents. Normative/prescriptive solutions are sought.

Problem Statement: Let H_0 and H_1 be two alternative hypotheses on the state of the environment and let there be N agents (sensors) who possess some stochastic information related to the state of the environment. In particular, we assume that each agent i observes a random variable y_i with known conditional distribution $P(y_i|H_j)$, $j = 0, 1$, given either hypothesis. We assume that all agents have information of the same quality, that is, the random variables are identically distributed. Each agent transmits a binary message to a central fusion center, based on his information y_i. The fusion center then takes into account all messages it has received to declare hypothesis H_0 or H_1 true. The problem consists of determining the optimal strategies of the agents as far as their choice of message is concerned. THis problem has been long recognized as a prototype problem in team decision theory: It is simple enough so that analysis may be feasible, but also rich enough to allow nontrivial insights into optimal team decision making under uncertainty.

Results: This being studied by Prof. J. Tsitsiklis and a graduate student, Mr. George Polychronopoulos. Under the assumption that the random variables y_i are conditionally independent (given either hypothesis), it is known that each agent should choose his message based on a likelihood ratio test. Nevertheless, we have constructed examples which show that even though there is a perfect symmetry in the problem, it is optimal to have different agents use different thresholds in their likelihood ratio tests. This is an unfortunate situation, because it severely complicates the numerical solution of the problem (that is, the explicit computation of the threshold of each agent). Still, we have shown that in the limit, as the number of agents becomes large, it is asymptotically optimal to have each agent use the same threshold. Furthermore, there is a simple effective computational procedure for evaluating this single optimal threshold.

We have also shown that if each agent is to transmit K-valued, as opposed to binary messages, then still each agent should use the same decision rule, when the number of agents is large. Unfortunately, however, the computation of this particular decision rule becomes increasingly broader as K increases.

We have investigated the case of M-ary (M > 2) hypothesis testing and constructed examples showing that it is better to have different agents use different decision rules, even in the limit as $N \to \infty$. Nevertheless, we have shown that the optimal set of decision rules is not completely arbitrary. In particular, it is optimal to partitition the set of agents into at most M(M-1)/2 groups and, for each group, each agent should use the same decision rule. The decision rule corresponding to each group and the proportion of the agents assigned to each group may be determined by solving a linear programming problem, at least in the case where the set of possible observations by each agent is finite.

In more recent work, the following have been accomplished.

(a) We studied the Neyman-Pearson (as opposed to Bayenian) version of the problem, in the case of M=2 hypothesis. The asymptotically optimal solution has been found and involves the Kullback-Liebler information distance.

(b) We considered a class of symmetic detection problems in which given any hypothesis H_i, each sensor has probability ε of making an observation indicating that some other hypothesis H_j is true. A simple numerical procedure has been found which completely solves this problem. Furthermore, a closed form formula for the optimal decision rules has been found for the case where the "noise intensity" ε is very small.

Future research will address the issue of the validity of asymptotic considerations when the number of agents N is moderate (N≈5) and will also investigate alternative (more complex) decision making architectures.

Documentation

[1] J. N. Tsitsiklis, "On Threshold Rules in Decentralized Detection," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986; also LIDS-P-1570, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, June 1986.

[2] J. N. Tsitsiklis, "Decentralized Detection by a Large Number of Sensors," LIDS-P-1662, April 1987; Submitted to *Mathematics of Control, Signals and Systems*.

3.3 COMMUNICATION REQUIREMENTS OF DIVISIONALIZED ORGANIZATIONS

Background: In typical organizations, the overall performance cannot be evaluated simply in terms of the performance of each subdivision, as there may be nontrivial coupling effects between distinct subdivisions. These couplings have to be taken explicitly into account; one way of doing so is to assign to the decisionmaker associated with the operation of each division a cost function which reflects the coupling of his own division with the remaining divisions. Still, there is some freedom in such a procedure: For any two divisions A and B it may be the responsibility of either decisionmaker A or decisionmaker B to ensure that the interaction does not deteriorate the performance of the organization. Of course, the decisionmaker in charge of those interactions needs to be informed about the actions of the other decisionmaker. This leads to the following problem. Given a divisionalized organization and an associated organizational cost function, assign cost functions to each division of the organization so that the following two goals are met: a) the costs due to the interaction between different divisions are fully accounted for by the subcosts of each division; b) the communication interface requirements between different divisions are small. In order to assess the communication requirements of a particular assignment of costs to divisions, we take the view that the decisionmakers may be modeled as boundedly rational individuals, that their decisionmaking process consists of a sequence of adjustments of their decisions in a direction of decreasing costs, while exchanging their tentative decisi... with other decisionmakers who have an interest in those decisions. We then require that there are enough communications so that this iterative process converges to an organizationally optimal set of decisions.

Problem Statement: Consider an organization with N divisions and an associated cost function $J(x_1,...,x_N)$, where x_i is the set of decisions taken at the i-th division. Alternatively, x_i may be viewed as the mode of operation of the i-th division. The objective is to have the organization operating at a set of decisions $(X_1,...,x_N)$ which are globally optimal, in the sense that they minimize the organizational cost J. We associate with each division a decisionmaker DM_i, who is in charge of adjusting the decision unables x_i. We model the decisionmakers as "boundedly rational" individuals; mathematically, this is translated to the assumption that each decisionmaker will slowly and iteratively adjust his decisions in a direction which reduces the organizational

costs. Furthermore, each decisionmaker does so based only on partial knowledge of the organizational cost, together with messages received from other decisionmakers.

Consider a partition $J(x_1,...,x_N) = \sum_{i=1}^{N} J^i(x_1,...,x_N)$ of the organizational cost. Each subcost J^i reflects the cost incurred to the i-th division and in principle should depend primarily on x_i and only on a few of the remaining x_j's. We then postulate that the decisionmakers adjust their decisions by means of the following process (algorithm):

(a) DM_i keeps a vector x with his estimates of the current decision x_k of the other decisionmakers; also a vector λ with estimates of $\lambda_i^k = \partial J^k/\partial x_i$, for $k \neq i$. (Notice that this partial derivative may be interpreted as DM_i's perception of how his decisions affect the costs incurred to the other divisions.

(b) Once in a while DMi updates his decision using the rule $x_i := x_i = \sum_{k=1}^{N} \lambda_i^k$, ($\gamma$ is a small positive scalar) which is just the usual gradient algorithm.

(c) Once in a while DM_i transmit his current decision to other decisionmakers.

(d) Other decisionmakers reply to DM_i, by sending an updated value of the partial derivative $\partial J^k/\partial x_i$.

It is not hard to see that for the above procedure to work it is not necessary that all DM's communicate to each other. In particular, if the subcost J_i depends only on x_i, for i, there would be no need for any communication whatsoever. The required communications are in fact determined by the sparsity structure of the Hessian matrix of the subcost functions J_i. Recall now that all that is given is the original cost function J; we therefore, have freedom in choosing the J_i's and we should be able to do this in a way that introduces minimal communication requirements; that is, we want to minimize the number of pairs of decisionmakers who need to communicate to each other.

The above problem is a prototype organizational design problem and we expect that it will lead to reasonable insights in good organizational structures. On the technical side, it may involve techniques and tools from graph theory. Once the above problem is understood and solved, the next step is to analyze communication requirements quantitatively. In particular, a distributed gradient algorithm such as the one introduced above converges only if the communication (between pairs of DM's who need to communicate) is frequent enough. We will then investigate the required frequencies of communication as a function of the strenght of coupling between different divisions.

Progress to Date: A graduatre student, C. Lee, supervised Prof. J. Tsitsiklis, has undertaken the task of task of formulating the problem of finding partitions that minimize the number of pairs of DM's who need to communicate to each other as the topic of his SM research. It was realized that with a naive formulation the optimal allocation of responsibilities, imposing minimal communication requirements, corresponds to the centralization of authority. Thus, in order to obtain more realistic and meaningful problems we are incorporating a constraint requiring that not agent should be overloaded. A number of results have been obtained for a class of combinatorial problems, corresponding to the problem of optimal organizational design, under limited communications. In particular certain special cases were solved and other special cases have been successfully reformulated as linear network flow of assignment problems, for which efficient algorithms are known. As simulation study is underway to validate the hypothesis that better task allocation results into better convergence.

Documentation: The Master's thesis of Mr. C. Lee will be ready by August 1987.

3.4 COMMUNICATION COMPLEXITY OF DISTRIBUTED CONVEX OPTIMIZATION

Background: The objective of this research effort is to quantify the minimal amount of information that has to be exchanged in an organization, subject to the requirement that a certain goal is accomplished, such as the minimization of an organizational cost function. The problem

becomes interesting and relevant under the assumption that no member of the organization "knows" the entire function being minimized, but rather each agent has knowledge of only a piece of the cost function. A normative/prescriptive solution is sought.

Problem Formulation: Let f and g be convex function of n variables. Suppose that each one of two agents (or decisionmakes) knows the function f (respectively g) , in the sense that he is able to compute instantly any quantities associated with this function. The two agents are to exchange a number of binary messages until they are able to determine a point x such that $f(g) + g(x)$ comes within ε of the minimum of f+g, where ε is some prespecified accuracy. The objective is to determine the minimum number of such messages that have to be exchanged, as a function of ε and to determine communication protocols which use no more messages than the minimum amount required.

Results: The problem is being studied by Professor John Tsitsiklis and a graduate student, Zhi-Quan Luo. We have shown that a least $0(n\log 1/\varepsilon)$ messages are needed and a suitable approximate and distributed implementation of ellispsoid-type algorithms work with $0(n^2\log^2 1/\varepsilon)$ messages. The challenge is to close this gap. This has been accomplished for the case of one-dimensional problem n=1, for which it has been shown that $0(\log 1/\varepsilon)$ messages are also sufficient. More recently, we have succeeded in generalizing the technique employed in the one-dimensional case, and we obtained an algorithm which is optimal, as far as the dependence of ε is concerned. The question of the dependence of the amount of communications on the dimension of the problem ($0(n)$ versus $0(n^2)$) seems to be a lot harder and, at present, there are no available techniques for handling it.

An interesting qualitative feature of the communication-optimal algorithms discovered thus far is the following: It is optimal to transmit aggregate information (the most significant bits of the gradient of the function optimized) in the beginning; then, as the optimum is approached more refined information should be transfered. This very intuitive result seems to correspond to realistic situtations in human decisionmaking. Another problem which is currently being

investigated concerns the case where are $K > 2$ decisionmakers cooperating for the minimization of $F_1 + ... + f_k$ where each f_i is again a convex function.

This problem turns out to be very hard, but some progress has been made on a simpler version of the problem. Namely, we considered the problem of evaluating a simple function (say the sume of K numbers) by a hierarchy (tree) of decisionmakers and tight bounds have been obtained on the required amount of communication.

Documentation:

[1] J. N. Tsitsiklis and Z.-Q. Luo, "Communication Complexity of Convex Optimization," LIDS-P-1617, Laboratory for Information and Decision Systems, MIT,October 1986; *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December, 1986; This paper has been submitted to the *Journal of Complexity*; also an invited talk was given at the *2nd Symposium on Complexity of Approximately Solved Problems*, Columbia University, New York, April 1987.

3.5 DISTRIBUTED ORGANIZATION DESIGN

Background: The bounded rationality of human decisionmakers and the complexities of the tasks they must perform mandate the formation of organizations. Organizational architectures distribute the decisionmaking workload among the members: different architectures impose different individual loads and result in different organizational performance. Two measures of organizational performance are accuracy and timeliness. The first measure of performance addresses in part the quality of the organization's response. The second measure reflects the fact that in tactical decisionmaking *when* a response is generated is also significant: the ability of an organization to carry out tasks in a timely manner is a determinant factor of effectiveness.

The scope of work was divided into three tasks:

(a) Evaluation of Alternative Organizational Architectures;
(b) Asynchronous Protocols; and
(c) Information Support Structures.

During this past year, the research effort was organized around three foci. In the first one, we continued to work on the development of analytical and algorithmic tools for the analysis and design of organizations. In the second, the focus was integration of the results obtained thus far through the development of a workstation for the design and analysis of alternative organizational architectures. Finally, an experimental program was initiated with the objective of collecting data necessary to calibrate the models and evaluate different architectures for distributed decisionmaking.

3.5.1 Design and Evaluation of Alternative Organizational Architectures.

In order to design an organization that meets some performance requirements, we need to be able to do the following:

(a) Articulate the requirements in qualitative and quantitative terms;
(b) Generate candidate architectures that meet some of the requirements;
(c) Evaluate the candidate organizations with respect to the remaining requirements;
(d) Modify the designs so as to improve the effectiveness of the organization;

The generalized Performance Workload locus has been used as the means for expressing both the requirements that the organization designer must meet and the performance characteristics of any specific design. Consider an organization with N decisionmakers. Then the Performance Workload space is an N+2 dimensional space in which two of the dimensions correspond to the measures of the organization's performance (say, accuracy and timeliness) and the remaining N dimensions correspond to the measure of the workload of each individual decisionmaker. Two loci can be defined. First, the Requirements locus is the set of points in this N+2 dimensional space that satisfy the performance and workload requirements associated with the task to be performed by the organization. The second, the System locus, is the set of points that are achievable by a particular design. The design problem can then be conceptualized as the reshaping and repositioning of the System locus in the Performance Workload space so that the requirements are met.

Several thesis projects were undertaken during this period. The individual problem statements and a decription of the progress to date follow:

Generation of Flexible Organizational Structures

Problem Statement: Develop a method for generating organizational forms that satisfy some structural and some application specific constraints.

Results: This problem has been addressed by P. Remy under the supervision of Dr. A. H. Levis. The first step in the procedure was the definition of the Petri Net and the corresponding data structure for the interacting decisionmaker. In the past, information sharing was allowed only between the situation assessment stage and the information fusion process. This assumption has been relaxed to allow four different forms of information sharing - each form depends on the source of the information (e.g., is one DM informing the other of his situation assessment or of his response?) and on the destination. For example, the situation assessment of one DM may be the input to the next one in a serial or hierarchical organization. After defining the set of possible interactions, a combinatorial problem could be formulated. The dimensionality of this problem is prohibitive, if no constraints on the structure are imposed. There are $2^{2n(2n-1)}$ organizational forms in this formulation, where n is the number of decisionmakers. These organizational forms are called Well Defined Nets (WDNs) of dimension n. An algorithmic approach has been developed that reduces the problem to a computationally tractable one.

A series of propositions, proved by Remy, set the theoretical basis of the algorithm. These propositions constitute significant extensions of Petri Net Theory. The first proposition establishes that if the source and the sink places of a Petri Net representing a WDN are combined into a single place and if the resulting Petri Net is strongly connected, then it is an event graph (a special class of Petri Nets).

Then, two sets of constraints are introduced to eliminate unrealistic organizational forms. The first set, structural constraints, define what kinds of interactions between decisionmakers must be ruled out. User-defined constraints allow the designer to introduce specific structural

characteristics that are appropriate (or are mandated) for the particular design problem.

The first structural constraint imposes a minimum degree of connectivity in the organization; it eliminates structures that do not represent a single integrated organization and ensures that the flow of information is continuous. The second constraint allows acyclical organizations only. This restriction is made to avoid deadlock and the circulation of messages. The third constraint prohibits one decisionmaker from sending the same data to different stages of another decisionmaker's model. This is a technical, model-specific restriction that recognizes the fact that the stages of decisionmaking are a modeling artifice that should not introduce extraneous complexity. The last constraint restricts the situation assessment stage to receiving a single input; multiple inputs can be received at the information fusion stage.

The user-defined constraints are arbitrary; they reduce the degrees of freedom in the design process. A WDN that satisfies the user-defined constraints is called an Admissible Organizational Form. An admissible form that also satisfies the structural constraints is a Feasible Organization.

The second proposition characterizes formally the admissible organizational forms as subsets of the set of WDNs. Furthermore, it introduces the concept of maximal and minimal elements of the sets. A maximal element of the set of Feasible Organizations is called a Maximally Connected Organization (MAXO) while a minimal one is called a Minimally Connected Organization (MINO).

The third proposition establishes that any feasible organization is bounded from above by at least one MAXO and from below by a least one MINO.

With this characterization of the feasible structures, what remains is to develop a procedure for generating them. The procedure is based on the concept of simple paths developed by Jin (or equivalently, on the s-invariants of Petri Net theory). The fourth and fifth propositions lead to the algorithm for generating feasible organizations. They show that one can construct the set of all the possible unions of simple paths. Then one can determine all the MAXOs and the MINOs of the set. These MAXOs and MINOs bound the solution set. Any feasible organization form is

a subset of a MAXO and has one or more MINOs as subsets. By adding simple paths to every MINO until a MAXO is reached, one can construct the complete set of Feasible Organizations.

This is a powerful result, both theoretically and computationally, that opens the way for generating classes of feasible organizational forms that meet, a priori, some structural and performance requirements. The partial ordering of the solutions (another result established by Remy) allows the use of lattice theory to analyze the properties of various architectures.

The work of Remy considered organizations with fixed structures: the decisions made by the organization members affected the processing of the task and resulted in different responses, but did not affect the structure of the organization. The next step is the consideration of flexible organizations in which the realized structure at any instant depends on the task that is being performed and the decisions being made. Jean Marc Monguillet, under the supervision of Dr. A. H. Levis, has began to investigate this question. At this point, the focus of the research is on understanding the meaning of the term "flexible architecture" and on the identification of appropriate mathematical tools for the description of such architectures.

Documentation:

[1] P. A. Remy, "On the Generation of Organizational Architectures Using PEtri Nets," LIDS-TH-1630, MS Thesis, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, December 1986.

[2] P.A. Remy, A.H. Levis, V.Y.-Y. Jin, "On the Design of Distributed Organizational Structures," *Proc. 10th IFAC World Congress*, Munich, FRG, July 1987; also accepted for publication in *Automatica*.

[3] P.A. Remy and A. H. Levis, "On the Generation of Organizational Architectures Using Petri Nets," LIDS-P-1634, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, December 1986; accepted for presentation at the *8th European Symposium on Applications and Theory of Petri Nets*, Zaragoza, Spain, 24-27 June, 1987.

Design of Organizations

Objective: Given a feasible organizational architecture, develop a methodology for (a) identifying the functions that must be performed by the organization in order that the task be accomplished,

(b) selecting the resources (human, hardware, software) that are required to implement these functions, and (c) integrating these resources - through interactions - so that the system operates effectively.

Progress to Date: This research problem is being investigated by Stamos K. Andreadakis under the supervision of Dr. A. H. Levis. The design methodology has been modified in order to address the following formulation of the design problem of decisionmaking organizations: Given a mission, design the DM organization that is accurate, timely, exhibits a task processing rate that is higher than the task arrival rate and whose decisionmakers are not overloaded. The design requirements explicitly stated are:

The accuracy J must be greater than a threshold J_0 or, equivalently, that the expected cost J be less than the threshold J_0:

$$J < J_0 \qquad [1]$$

The timeliness measure T be less than a threshold T_0:

$$T < T_0 \qquad [2]$$

The task processing rate R be greater than the task arrival rate R_0:

$$R > R_0 \qquad [3]$$

The constraints that must be observed are: each decisionmaker must not be overloaded, i.e., a decisionmakers' information processing rate F be less than the rationality threshold F_0:

$$F < F_0 \qquad [4]$$

The proposed design methodology has two stages:

In the first stage the Petri Net of the data flow is constructed. Each function is represented by a transition, while the associated data (information) and constraints are represented by places. This Petri Net depicts the flow of information from function to function, as well as the parallel

(concurrent) and serial operations.

Since a multitude of data flow architectures can be constructed for a particular mission, it is necessary to classify them in order to keep the design problem tractable. Current research focuses on the classification of data flow architectures and on the selection of a representative structure for each data flow class.

The objective of the first stage is to compute an estimate of the processing rate range and to select the number of processing channels in order to satisfy the processing rate requirement. In order to satisfy the workload constraints, the total activity of each function in the Petri Net is computed, using Information Theory. Then a representative (average) value for the information processing capacity of the human is selected and the expected execution time of each function is computed by dividing the total activity by the processing capacity. In all subsequent calculations of the information processing rate, response time, and timeliness measures, these processing time estimates are used. Thus, the workload constraints will always be satisfied.

The maximum and minimum processing rates of the organization can be computed as follows:

The processing rate of each transition is computed by dividing the information processing capacity by the total activity of the function that is represented by the transition. Assuming that each transition is assigned to a different decisionmaker, the maximum processing rate of the Decision Making Organization (DMO) being designed is equal to the minimum of the rates of the individual transitions.

An estimate of the minimum processing rate is obtained as follows:

The expected processing time corresponding to each information flow path is computed by summing the expected processing times of its transitions. The rate estimate is the inverse of the maximum expected processing time. The actual minimum rate can be even lower due to communications delays.

If the processing rate is smaller than the input rate, multiple processing channels, which are

copies of the basic information flow net, must be introduced so that the arriving tasks can be assigned to alternate processing channels of the DMO. The number of the alternate channels is computed by dividing the input rate by the processing rate and rounding up to an integer value. Consequently, the processing rate requirements will be satisfied.

In the second stage, the Petri Net depicting the information processing is augmented and is transformed into the Petri Net of the DMO. This Petri Net delimits the functions performed by each decisionmaker by introducing resource places that represent the constraints on the decisionmakers to perform one function at any time. Each of these places is connected so that it is the output of the last and the input of the first transition allocated to the decisionmaker.

This Petri Net also depicts the communications among members of the DMO by representing each communication process by a transition and the respective protocols using the appropriate places and connectors. When allocating functions to decisionmakers the following sets of constraints must be met:

The functions allocated to any decisionmaker must observe the input-output relationships imposed by the Petri Net and must process data pertinent to the same subtask of the DM organization. They must also belong to different time zones or slices of the Petri Net, i.e., they must process data at different times.

Transitions belonging to an information flow path observe the input-output relationships and are executed sequentially. Thus they satisfy both of the above constraints. Transitions on different flow paths violate the first requirement (input-output). Thus the feasible solutions to the function allocation problem are the sets of functions that correspond to the information flow paths. Consequently, the functions of each information flow path are allocated to a different decisionmaker.

In general, the functions represented by the transitions require different specialization from the decisionmakers. Hence when considering alternate sets of functions, the respective training requirements must be considered. Due to specialization requirements, it may be necessary to allocate some of the transitions of an information flow path to one decisionmaker and the

remaining to another.

Another consideration is the sensitivity of the timeliness measure to communications jamming. Consequently, function allocation resulting in having two or more communication links on any information flow path is more sensitive to communications jamming than that resulting in only one communication link on an information flow path.

To evaluate a design, the accuracy of response, the expected response time, and the processing rate of the DMO are computed for all decision strategies. Then a Measure of Effectiveness of the DMO is defined in the strategy space as the ratio of decision strategies that satisfy the requirements to the total number of decision strategies. If the MOE value is satisfactory then the design is accepted; else the design is modified until a satisfactory MOE value is obtained.

Documentation:

[1] A. H. Levis and S. K. Andreadakis, "Computer-Aided Analysis of Organizations," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.

[2] S.K. Andreadakis and A. H. Levis, "Accuracy and Timeliness in Decision-Making Organizations," *Proc. 10th IFAC World Congress*, July 27-31, 1987, Munich, FRG.

Presentations:

[1] S. K. Andreadakis and A. H. Levis, Accuracy and Timeliness in Decision-Making Organizations," *9th MIT/ONR Workshop on C^3 Systems*, June 1986, Monterey, California.

[2] S. K. Andreadakis and A. H. Levis, "Design Methodology for Decision-Making Organizations," *C^3 Symposium*, June 1987, Washington DC.

The ability of a distributed tactical decisionmaking organization to carry out its tasks in a timely manner depends on two types of constraints. The first type is related to the internal organizational structure that determines how the various operations occur in the process: some tasks are processed sequentially, while others are processed concurrently. The sequential and concurrent events are coordinated by the communication and execution protocols among the individual organization members.

The second type includes time and resource constraints. The time constraints derive from the task execution times -- the time necessary to perform each task. The organization also has limited resources; depending on which of the resources are available at a given instant, some activities can take place while others must be delayed. The Petri Net formalism provides a convenient tool for analyzing the behavior of organizations with asynchronous protocols that allow for concurrent processing.

Performance Evaluation of Organizations with Decision Aids

In the work of Remy and Andreadakis, the organizations consist of humans alone. Alternatively, the effect of decision aids is subsumed in the model of the organization member. However, in considering command and control systems, it is necessary that the contributions and effects of decision aids be made explicit. Jean Louis Grevet, under the supervision of Dr. A. H. Levis, has started a thesis project that will try to build on the case study done by S. Weingaertner in his thesis and develop a methodology for the modeling and analysis of decision aids in an organization.

Performance Evaluation of Organizations with Asynchronous Protocols

Problem Statement: In earlier work by Jin, the response time of a decisionmaking organization was computed using an algorithm based on the Petri Net representation. The definition of response time was the time interval between the moment a stimulus is received by the organization and the moment a response is made. This measure of performance is a static measure insofar that it assumes that there are no other tasks being processed by the organization. A more realistic estimate of response time will be obtained, if the dynamic behavior of the organization is taken into account. More precisely, the research problem is to evaluate the performance of DMO with respect to the following time-related measures:

(a) Maximum Throughput Rate: This is the maximum rate at which external inputs can be processed; a higher rate would lead to the formation of queues of unbounded length.

(b) Execution Schedule: Let processing of arriving inputs start at t_0 and let the inputs be

processed at the maximum throughput rate. The earliest isntants of time at which the various tasks can be performed in the repetitive process constitute the optimum execution schedule; any other schedule will lead to longer response times.

Results: The time-related performance of a DMO, as measured by the maximum throughput rate and the execution schedule, has been analyzed and evaluated by Herve P. Hillion under the supervision of Dr. A. H. Levis. The approach was based on modeling the DMO as a Timed Petri Net. Two constraints have been modeled to characterize the bounded rationality of human decisionmakers. The time associated with individual processes reflects a processing rate limitation, while the resource limitation models the limited capacity of short-term memory, which bounds the amount of information that a DM can handle at the same time. Both considerations are modeled as a constraint on the total number of inputs that can be processed simultaneously.

The maximum throughput rate has been expressed as a function of the resource and time constraints in the following manner: The inclusion of the resource constraints in the Petri Net model results in directed circuits (or loops) which are characterized by: (a) the circuit processing time, μ, defined as the sum of the different task processing times of the circuit. μ represents the amount of time it takes one input to complete the processing operations of the circuit; (b) the resources available, n, which bound the total number of inputs that can be processed at the same time in the circuit.

For a given circuit, the ratio n/μ characterizes the average circuit processing rate. The minimum average circuit processing rate, taken over all the directed circuits of the net, determines the maximum throughput rate of the deterministic systems, i.e., when all the task processing times are deterministic. For the case of stochastic processing times, an upper bound is obtained for the maximum throughput rate. In that case, the average circuit processing time can be computed. The determination of the critical circuits, for which the corresponding average processing rate is minimal, provides a clear way of comparing different organizations. These critical circuits are the ones that, because of the time and resource constraints, bound the throughput rate. Therefore, there is now a direct way to identify how different constraints affect organizational performance. Consequently, the problem of modifying the right constraints so as to improve the performance

of the organizations (and meet mission requirements) becomes transparent.

A method for obtaining and analyzing the exact execution schedule when processing times are deterministic has been developed. A representation, defined by the slices of the Petri Net, allows for the precise characterization of the causal relations in the DMO. The causal relationships result in the partial ordering of the different operations. The execution schedule so obtained determines the earliest instants at which the various tasks can be executed in real-time for a process that occurs repetitively.

The contribution of this research is that it develops two MOPs that characterize the time-related behavior of a distributed tactical decisionmaking organization. Furthermore, the concepts and algorithms developed are oriented toward design: they indicate which design parameters need to be changed to meet requirements.

Documentation:

[1] H. P. Hillion, "Performance Evaluation of Decisionmaking Organizations Using Timed Petri Nets," LIDS-TH-1590, MS Thesis, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, August 1986.

[2] H. P. Hillion and A. H. Levis, "Timed Event-Graph and Performance Evaluation of Systems," LIDS-P-1639, Laboratory for Information and Decision Systems, MIT, January 1987; accepted for presentation at the *8th European Workshop on Applications and Theory of Petri Nets*, Zaragoza Spain, 24-27 June 1987.

3.5.2 Computer Aided Evaluation of System Architectures

During the last few years, a number of problems regarding distributed tactical decisionmaking have been addressed and models, algorithms, and methods have resulted that are useful for answering specific aspects of the overall problem. In order to integrate these results into a consistent methodology and to provide the means for designing an experimental program, a computer aided design system has been developed. While the primary support for this development has been by the Basic Research Group (BRG) of the Technical Panel on C^3 of the Joint Directors of Laboratories, there has been sufficient contribution by the staff of this project

to warrant its inclusion in this report. The components of the system contributed by this project are identified by "DTDM support" in the detailed description that follows.

The design workstation has been named **CAESAR** for Computer Aided Evaluation of System ARchitectures. It consists of four major components:

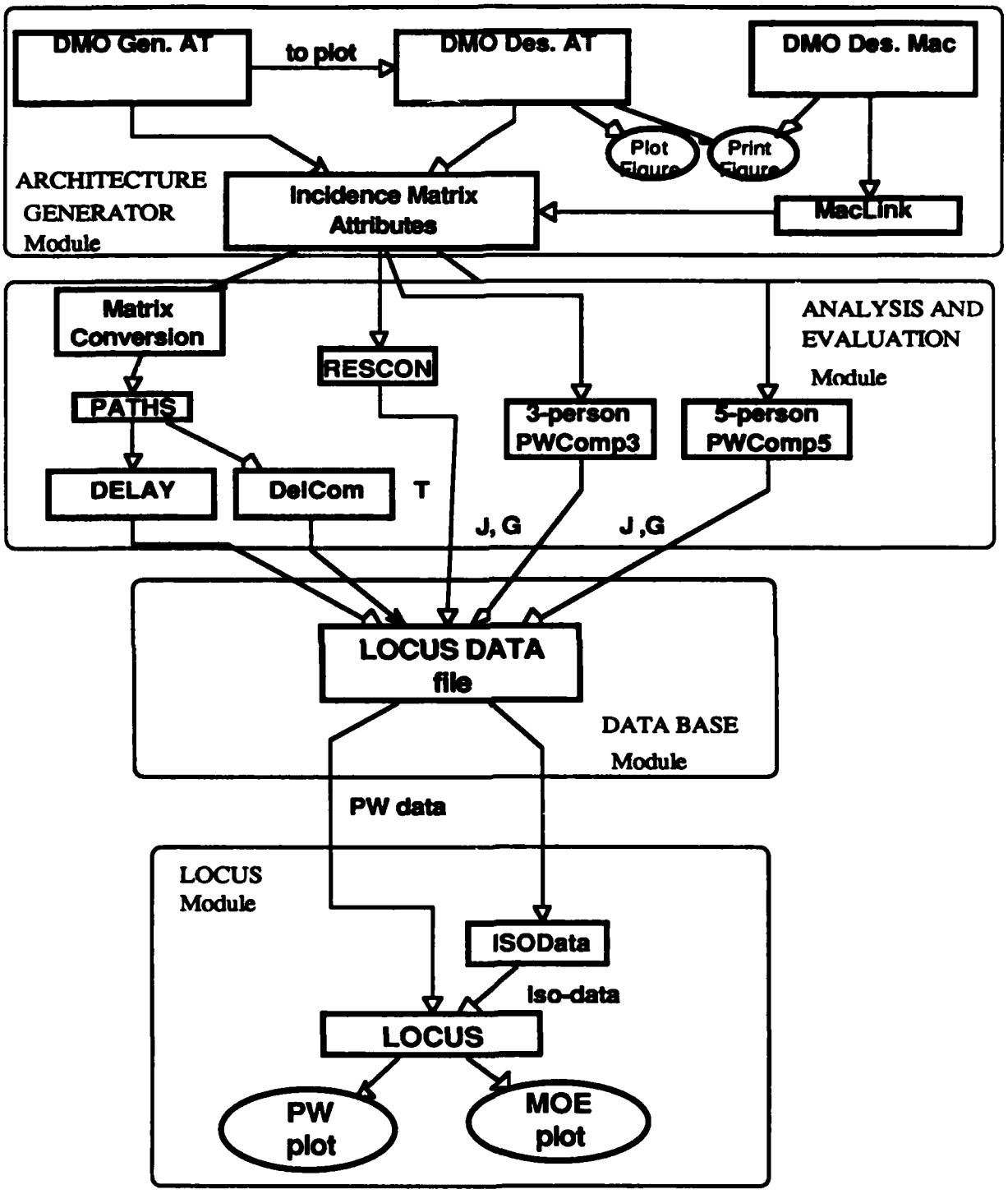
The **Architecture Generator** which constructs feasible organizational forms using Petri Nets.

The **Analysis and Evaluation Module** which contains many of the algorithms for the computation of the Measures of Performance.

A **Data Base** which is used to store the results of the analysis.

The **Locus Module** that constructs the generalized Performance Workload locus of an organization and can be used to evaluate Measures of Effectiveness.

The structure of the software system is shown in Figure 1. The individual modules and their status are described below.



CAESAR· Computer - Aided Evaluation of System ARchitectures v. 1.0. 3/87

LIST OF MODULES IN CAESAR

A. ARCHITECTURE GENERATOR

DMO Gen.AT

Program that generates the Petri Nets of Decisionmaking Organizations that satisfy a set of structural constraints, as well as constraints imposed by the user. The algorithm is based on P. Remy's thesis (1986) and has been implemented in DOS 3.0 © IBM, using Turbo Pascal 3.01A ©Borland International and Screen Sculptor ©Software Bottling Company.
Status: Program operational. It requires an interface with DMO Des.AT so that a graphical description of the feasible architectures can be obtained directly. (DTDM support)

DMO Des.AT

Interactive graphics program for the construction of the Petri Nets of arbitrary organizational architectures. It can be used to create and store subsytems and to combine them to form large organizational structures. Program, developed by I. Kyratzoglou, also creates the analytical description of the Petri Nets. Implemented in DOS 3.0, Professional Fortran, Graphics Tool Kit, and Graphic Kernel System, *all ©IBM.*
Status: To be completed by June 1 (JDL support).

DMO Des.Mac

Interactive graphics program for the construction of the Petri Nets of arbitrary organizations. It can be used to design organizations of arbitrary size through the use of nested subnets. Program developed for the Apple Macintosh by the Meta Software Corp. using the Design Open Architecture System ©Meta Software Corp. The program creates the analytical description of the Petri Net, as well as store functions and attributes represented by the transitions, places, and connectors. Program enhanced by J. L. Grevet to be consistent with analytical description of Petri Nets used in various algorithms.
Status: Program operational. (JDL support)

MacLink ©Dataviz

Commercial software for for converting and transmitting files between the DOS machines and the Macintosh.
Status: MacLink has been installed and is operational: it can transfer the data structure of a Petri Net from the DMO Des.Mac module to the Analysis and Evaluation Module.

Incidence Matrix / Attributes

Standard form for the data structure of Petri Nets. The files contain the incidence matrix or flow matrix of the Petri Net and the attributes and functions associated with the elements of the net.
Status: Standard version of incidence matrix has been implemented; the specifications for the attribute file are being developed. Expected completion date: July 1. (DTDM support)

B. ANALYSIS AND EVALUATION MODULE

Matrix Conversion

Simple algorithm that transforms the incidence matrix into the interconnection matrix used in Jin's algorithm. Algorithm in Turbo Pascal 3.01A.
Status: Algorithm developed by Jin is operational. (DTDM support)

Paths

Algorithm developed by Jin in her thesis that determines all the simple paths and then constructs the concurrent paths in an organizational architecture. This is an efficient algorithm that obtains the answers by scanning the interconnection matrix.Algorithm in Turbo Pascal 3.01A.
Status: Program is operational. (DTDM support)

Delay

Simple algorithm that calculates path delays and expected delay when processing delays are constant. Algorithm in Turbo Pascal 3.01A.
Status: Algorithm is operational. (DTDM support)

Del Com

Algorithm developed by Andreadakis that calculates measures of timeliness when the processing delays are described by beta distributions. It also accounts for the presence of jamming and its effect on timeliness. Algorithm in Turbo Pascal 3.01A.
Status: Problem specific version operational; general version to be completed by September 1. (DTDM support)

Res Con

Algorithm developed by Hillion in his thesis that calculates the maximun throughput in a Timed Event Graph, a special class of Petri Nets. It also detrmines the optimal schedule in the presence of resource and time constraints. The procedure incorporates an algorithm proposed by Martinez and Silva for determining simple paths through the calculation of s-invariants.
Status: Independent version of algorithm is operational;

integrated version in workstation to be operational by June 1. (JDL support)

PW Comp 3

Algorithm for the computation of a three-person organization's performance measure J (Accuracy) and the workload of each one of the decisionmakers. The algorithm computes the accuracy of the response and the workload for each admissible decision strategy. This version was developed by Andreadakis in Turbo Pascal.
Status: Program is operational. (DTDM support)

PW Comp 5

A variant of PW Comp 3, but for a five-person organization modeling the ship control party of a submarine. Algorithm developed by Weingaertner as part of his thesis. Implemented in Turbo Pascal.
Status: Program is operational.

C. DATA BASE MODULE

LOCUS Data File

Data file in which the results from the evaluation of a decisionmaking organization are stored. The file, as currently structured, can accommodate five measures of performance - accuracy, timeliness, and workload for three persons. It also contains four indices that specify the decision strategy associated with each record.
Status: Three-person organization version operational. General structure to be implemented by June 1. (JDL support)

D. LOCUS MODULE

LOCUS

Graphics plotting program that generates two or three dimensional loci or two- and three-dimensional projections of higher dimensional loci. This is the basic program used to construct the Performance - Workload locus of an organization. Basic version developed by Andreadakis and Bohner and described in latter's thesis.
Status: Version using professional graphics controller is operational. Revised transportable version adhering to the VDI standard and with improved user interface is also operational. (DTDM support)

ISO Data

Algorithm for obtaining some measures of effectiveness from the measures of performance stored in the Locus Data file. Specifically, it finds isoquants: e. g., locus of constant accuracy, or constant workload.
Status: New version for microcomputers being

implemented by J. Azzola using a design by Weingaertner. Expected date of completion is June 1.

E. INPUT / OUTPUT

Output: By adopting the Virtual Device Interface (VDI) standard and the Enhanced Graphics standard, it is possible to develop a version of the CAESAR software that is transportable to other IBM PC ATs or compatibles and to drive a wide variety of output devices: various monitors, printers, laser printers, and pen plotters.

Input: A uniform user interface with windowing capability is needed to make the system useable by analysts and designers. Commercially available software are being investigated to select the most appropriate one. Expected completion date is September 1.

We expect to have the transportable version of CAESAR operational by September 1, 1987 and demonstarate it at the next annual review of the DTDM program. There have been some delays primarily due to a five month delay in obtaining a properly configured AT compatible machine that serves as the workstation.

3.5.3 Design of Experiments

A major application of CAESAR is in the design and analysis of experiments in which different organizational forms will be evaluated. At this time, V. Jin has initiated a project, under the supervision of Dr. A. H. Levis, in which she is assessing the applicability of certain methodologies in the physical sciences for the design of experiments to the behavioral sciences. In the meantime, with funding from Joint Directors of Laboratories, an experiment is being carried out to determine the stability of the bounded rationality constraint and to obtain, if possible, values for it.

5. RESEARCH PERSONNEL

Prof. Michael Athans,	Co-principal investigator
Dr. Alexander H. Levis,	Co-principal investigator
Prof. John Tsitsiklis,	
Dr. Jeff T. Casey	
Mr. Stamatios Andreadakis	graduate research assistant (Ph.D)
Ms. Victoria Jin	graduate research assistant (Ph.D)
Mr. Jason Papastavrou	graduate research assistant (Ph.D)
Ms. Chongwan Lee	graduate research assistant (M.S.)
Mr. Jean-Louis Grevet	graduate research assistant (M.S.)
Mr. Jean-Marc Monguillet	graduate research assistant (M.S.)

6. DOCUMENTATION

6.1 Theses

[1] J. N. Tsitsiklis, "Problems in Decentralized Decisionmaking and Computation," Ph.D. Thesis, Dept. of EECS also, Report LIDS-TH-1424, Laboratory for Information and Decision Systems, MIT, December 1984.

[2] G. J. Bejjani, "Information Storage and Access in Decisionmaking Organizations," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1434, Laboratory for Information and Decision Systems, MIT, January 1985.

[3] Y. V.-Y. Jin, "Delays for Distributed Decisionmaking Organizations," MS Thesis, Mechanical Engineering, also, Report LIDS-TH-1459, Laboratory for Information and Decision Systems, MIT, May 1985.

[4] K. L. Boettcher, "A Methodology for the Analysis and Design of Human Information Processing Organizations," Ph.D., Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH-1501, Laboratory for Information and Decision Systems, MIT, September 1985.

[5] R. P. Wiley, "Performance Analysis of Stochastic TImed Petri Nets," Ph.D. Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH---1525, Laboratory for Information and Decision Systems, MIT, January 1986.

[6] J. D. Papastavrou, "Distributed Detection with Selective Communications," MS Thesis, Dept. of Electrical Engineering and Computer Sciences, also, Report LIDS-TH-1563, Laboratory for Information and Decision Systems, MIT, May 1986.

[7] C.M. Bohner, "Computer Graphics for Systems Effectiveness Analysis," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1573, Laboratory for Information and Decision Systems, MIT, June 1986.

[8] H. P. Hillion, "Performance Evaluation of Decisionmaking Organizations Using Timed Petri Nets," MS Thesis, Technology and Policy Program, also, Report LIDS-TH-1590, MIT, August 1986.

[9] P. A. Remy, "On the Generation of Organizational Architectures Using Petri Nets," MS Thesis, Technology and Policy Program, also, Report LIDS-TH--1630, Laboratory for Information and Decision Systems, MIT, December 1986.

6.2 Technical Papers

[1] R. P. Wiley and R. R. Tenney, "Performance Evaluation of Stochastic Timed Decision-Free Petri Nets," *Proc. 24th IEEE Conference on Decision and Control,* December 11-13, 1985, Ft. Lauderdale, FL, pp. 558-563.

[2] G. J. Bejjani and A. H. Levis, "Information Storage and Access in Decisionmaking Organizations," LIDS-P-1466, Laboratory for Information and Decision Systems, MIT, May 1986.

[3] J. N. Tsitsiklis and M. Athans, "On the Complexity of Decentralized Decisionmaking and Detection Problems, *IEEE Trans. on Automatic Control*, Vol. AC-30, No. 5, May 1985.

[4] V. Y.-Y. Jin and A. H. Levis, "Computation of Delays in Acyclical Distributed Decisionmaking Organizations," LIDS-P-1488, Laboratory for Information and Decision Systems, MIT, August 1985.

[5] K. L. Boettcher, and R. R. Tenney, "A Case Study in Human Team Decisionmaking," Proc. Systems, Man, and Cybernetics, Tucson, AZ, November 1985.

[6] K. L. Boettcher and R. R. Tenney, "On the Analysis and Design of Human Information Processing Organizations," LIDS-P-1503, *Proc. 8th MIT/ONR Workshop on C^3 Systems*, December 1985.

[7] K. L. Boettcher and R. R. Tenney, "Distributed Decisionmaking with Constrained Decisionmakers - A Case Study," *Proc. 8th MIT/ONR Workshop on C^3 Systems*, December 1985, also in *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-16, No. 6, November/December 1986.

[8] R. P. Wiley and R. R. Tenney, "Calculating TImed-Related Performance Measures of a Distributed Tactical Decisionmaking Organization Using Stochastic Timed Petri Nets,"*Proc. 8th MIT/ONR Workshop on C^3 Systems*, December 1985.

[9] V. Y.-Y. Jin, A. H. Levis, and P. A. Remy, "Delays in Acyclical Distributed Decisionmaking Organziations," LIDS-P-1528, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, also in *Proc. 4th IFAC/IFORS Symposium on Large Scale Systems: Theory and Applications*, Zurich, Switzerland, August 1986.

[10] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Trans. on Automatic Control*, Vol. AC-31, No. 9, September 1986.

[11] J. Papastravrou and M. Athans, "A Distributed Hypotheses-Testing Team Decision Problem with Communications Cost," LIDS-P-1538, Laboratory for Information and Decision Systems, MIT, February 1986, also in *Proc. 25th IEEE Conference on Decision and Control,* Athens, Greece, December 1986.

[12] J. Tsitsiklis, "On Optimal Thresholds in Decentralized Detection," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.

[13] A. H. Levis and S. K. Andreadakis, "Computer-Aided Design of Organizations," *Proc. 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986.

[14] P. Remy, A. H. Levis, V. Jin, "On the Design of Distributed Organizational Structures," LIDS-P-1581, Laboratory for Information and Decision Systems, MIT, July 1986; also *Proc. 10th IFAC World Congress*, Munich, FRG, July 1987. To appear in *Automatica.*

[15] M. Athans, "Command-and-Control (C^2) Theory: A Challenge to Control Science," LIDS-P-1584, Laboratory for Information and Decision Systems, MIT, September 1986; also, *IEEE Trans. on Automatic Control*, April 1987.

[16] P.A. Remy and A. H. Levis, "On the Generation of Organizational Architectures Using Petri Nets," *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987.

[17] H.P. Hillion and A.H. Levis, "Timed Event-Graph and Performance Evaluation of Systems," *Proc. Eighth European Workshop on Applications and Theory of Petri Nets*, Zaragoza, Spain, June 24-26, 1987.

April 1987

DECENTRALIZED DETECTION BY A LARGE NUMBER OF SENSORS[1]

John N. Tsitsiklis[2]

ABSTRACT

We consider the decentralized detection problem, in which a number N of identical sensors transmit a finite–valued function of their observations to a fusion center which then decides which one of M alternative hypotheses is true. We consider the case where the number of sensors tends to infinity. We then show that it is asymptotically optimal to divide the sensors into $M(M-1)/2$ groups, with all sensors in each group using the same decision rule in deciding what to transmit. We also show how the optimal number of sensors in each group may be determined by solving a mathematical programming problem. For the special case of two hypotheses and binary messages the solution simplifies considerably: it is optimal (asymptotically, as $N \to \infty$) to have each sensor perform an identical likelihood ratio test and the optimal threshold is very easy to determine numerically.

1. Research supported by the Army Research Office under grant DAAL03–86–K–0171 and by the Office of Naval Research under contract N00014–84–K–0519.

2. Room 35–214, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

I. INTRODUCTION AND PROBLEM DEFINITION.

The (static) decentralized detection problem is defined as follows. There are M hypotheses $H_1, \cdots, H_M$, with known prior probabilities $P(H_i) > 0$ and N sensors. Let Y be a set endowed with a σ–field $\mathcal{F}$ of measurable sets. Let y_i, $i = 1, ..., N$, the observation of the i-th sensor, be a random variable taking values in Y. We assume that the y_i's are conditionally independent and identically distributed, given either hypothesis, with a known conditional distribution $P(y|H_j)$, $j = 1, ..., M$. Let D be a positive integer. Each sensor i evaluates a D–valued message $u_i \in \{1, ..., D\}$, as a function of its own observation; that is $u_i = \gamma_i(y_i)$, where the function $\gamma_i : Y \mapsto \{1, ..., D\}$ is the decision rule of sensor i and is assumed to be a measurable function. The messages $u_1, ..., u_N$ are all transmitted to a fusion center which declares one of the hypotheses to be true, based on a decision rule $\gamma_0 : \{1, ..., D\}^N \mapsto \{1, ..., M\}$. That is, the final decision u_0 of the fusion center is given by $u_0 = \gamma_0(u_1, ..., u_N)$. The objective is to choose the decision rules $\gamma_0, \gamma_1, ..., \gamma_N$ of the sensors and the fusion center so as to minimize the probability of error in the decision of the fusion center. (An alternative formulation of the problem, of the Neyman–Pearson type will be also considered in the last section.)

The above defined problem and its variants have been the subject of a fair amount of recent research [TS, E, TA, LS], especially for the case of binary hypotheses ($M = 2$) and binary messages ($D = 2$). For the latter case, it is known that any optimal set of decision rules has the following structure. Each one of the sensors evaluates its message u_i using a likelihood ratio test with an appropriate threshold. Then, the fusion center makes its decision by performing a final likelihood ratio test. (Here, the messages received by the center play the role of its observations.) Without the conditional independence assumption we introduced, this result fails to hold and the problem is intractable (NP-hard), even for the case of two sensors [TA]. Assuming conditional independence, the optimal value of the threshold of each sensor may be obtained by finding all solutions of a set of coupled algebraic equations (which are the person–to–person optimality conditions for this problem) and by selecting the solution which results to least cost. Unfortunately (and contrary to intuition), even if the observations of each sensor are identically distributed (given either hypothesis) it is

not true that all sensors should use the same threshold (see the Appendix for an example). This renders the computation of the optimal thresholds intractable, when the number of sensors is large. To justify this last claim, consider what is involved in just evaluating the cost associated to a fixed set $\gamma_0, \gamma_1, ..., \gamma_N$, of decision rules if each sensor uses a different threshold. In order to evaluate the expected cost, we have to perform a summation over all possible values of $(u_1, ..., u_N)$, which means that there are 2^N terms to be summed. (This is in contrast to the case of equal thresholds in which the u_i's are identically distributed and therefore the binomial formula may be used to obtain a sum with only $N+1$ summands.) Of course, to determine an optimal set of decision rules this effort may have to be repeated a number of times. This suggests that the computational effort grows exponentially with the number N of sensors.

The above discussion motivates the main results of this paper which show that, for the case $M = 2$, $D = 2$, it is asymptotically optimal to have each sensor use the same threshold and provides a simple method for computing the optimal threshold. For the general case of $M > 2$ hypotheses, it is no longer true, not even in the limit as $\Lambda \to \infty$, that each sensor should use the same decision rule. Nevertheless, we show that, as $N \to \infty$, at most $M(M-1)/2$ different decision rules need to be used by the sensors. The determination of an asymptotically optimal set of decision rules is still a hard computational problem, except for the case where the observation set Y is finite and of small cardinality.

Notation: Throughout, P_i will stand for the (conditional) measure $P(\cdot|H_i)$ on $(Y, \mathcal{F})$, under hypothesis H_i. Furthermore, $E_i[\cdot]$ will stand for expectation with respect to the measure P_i.

II. THE BAYESIAN PROBLEM.

We start by noticing that, having fixed the decision rules $\gamma_1, ..., \gamma_N$ of the sensors, the optimal decision for the fusion center is determined by using the maximum a posteriori probability (MAP) rule. (The messages to the fusion center are to be thought as measurements available to it.) Thus, γ_0 is straightforward to determine in terms of $\gamma_1, ..., \gamma_N$. For this reason, we shall be concerned only with the optimization with respect to $(\gamma_1, ..., \gamma_N)$. Any such set of decison rules will be denoted,

for convenience, by γ^N.

We introduce some more notation. Let Γ be a set of decision rules among which the decision rules of each sensor are to be selected. In general, we should take Γ to be the set of all (measurable) functions from Y into the set $\{1,...,D\}$. However, we may, for some reason, wish to restrict to a smaller class of decision rules, possibly having some simplifying structure. We return to this issue in Section III. Let Γ^N be the Cartesian product of Γ with itself, N times. For any $\gamma^N \in \Gamma^N$, let $J_N(\gamma^N)$ be the probability of an erroneous final decision by the fusion center (always assuming that the fusion center uses the MAP rule). We are concerned with the minimization of $J_N(\gamma^N)$, over all $\gamma^N \in \Gamma^N$, when N is very large.

It is easy to show that, as the number of sensors grows to infinity, the probability of error goes to zero, for any reasonable set of decision rules, in fact exponentially fast. Consequently, we need a more refined way of comparing different sets of decision rules, as $N \to \infty$. To this effect, for any given value of N and any set γ^N of decision rules for the N–sensor problem, we consider the exponent of the error probability defined by

$$r_N(\gamma^N) = \frac{\log J_N(\gamma^N)}{N}.$$

Let $R_N = \inf_{\gamma^N \in \Gamma^N} r_N(\gamma^N)$ be the optimal exponent. Let Γ_0^N be the set of all $\gamma^N \in \Gamma^N$ with the property that the set $\{\gamma_1,...,\gamma_N\}$ has at most $M(M-1)/2$ different elements. Let $Q_N = \inf_{\gamma^N \in \Gamma_0^N} r_N(\gamma^N)$ be the optimal exponent, when we restrict to sets of decision rules in Γ_0^N. The following result shows that, asymptotically, optimality is not lost, if we restrict to Γ_0^N.

Theorem 1: Subject to Assumption 1 below, $\lim_{N\to\infty}(Q_N - R_N) = 0$.

The rest of this section is devoted to the proof of Theorem 1. We first need to introduce some auxiliary tools.

Let us fix some $\gamma \in \Gamma$. The mapping from the true hypothesis H_i to the decision of a sensor employing the decision rule γ may be thought of as a noisy channel which is completely described by the probabilities

$$p_i^\gamma(d) = P_i(\gamma(y) = d).$$

The ability of such a channel to discriminate between hypotheses H_i and H_j $(i \neq j)$ may be quantified by a function $\mu_{ij}(\gamma, s)$, $s \in [0,1]$, defined by the following formula [SGB]:

$$\mu_{ij}(\gamma, s) = \log \left[\sum_{d=1}^{D} (p_i^\gamma(d))^{1-s} (p_j^\gamma(d))^s \right]. \tag{1}$$

We use here the convention $0^0 = 0$; thus, the summation in (1) is to be performed only over those d's for which $p_i^\gamma(d) p_j^\gamma(d) \neq 0$. Assuming that $\mu_{ij}(\gamma, s)$ is not infinite, it is easy to see that $\mu_{ij}(\gamma, s)$, is infinitely differentiable, as a function of s, and its derivatives are continuous on $[0,1]$, provided that we define the derivative at an endpoint as the limit when we approach the endpoint from the interior.

Notice that, for any fixed γ, the function $\mu_{ij}(\gamma, s)$ is equal to $E[e^{sX}]$, where X is the log–likelihood ratio of the distributions $p_j^\gamma(\cdot)$ and $p_i^\gamma(\cdot)$, where the expectation is with respect to the distribution $p_i^\gamma(\cdot)$. As is well–known, minimizing the characteristic function of a random variable X yields tight bounds on the probability of large deviations of X from its mean. Since in this case X is the log–likelihood ratio, this method leads to tight bounds on the probability of error. One particular such result that we will use is taken from [SGB]:

Lemma 1: Let there be two hypotheses H' and H''. Let $x_1, ..., x_N$ be measurements taking values in a finite set $\{1, ..., D\}$, which are conditionally independent given the true hypothesis and suppose that the conditional distribution of x_i, when H is true, is described by $p_H^i(d) = P(x_i = d \mid H)$. Let

$$\mu(i, s) = \log \left[\sum_{d=1}^{D} (p_{H'}^i(d))^{1-s} (p_{H''}^i(d))^s \right]$$

and $\mu(s) = \sum_{i=1}^{N} \mu(i, s)$. Assume that $\mu(i, s)$, $\mu'(i, s)$, $\mu''(i, s)$ exist and are finite, where a prime stands for differentiation with respect to s. Let s^* minimize $\mu(s)$, over $s \in [0,1]$. Then,

a) There exists a decision rule for deciding between H' and H'', on the basis of the measurements $x_1, ..., x_N$, for which

$$P(\text{decide } H' \mid H'' \text{ is true}) + P(\text{decide } H'' \mid H' \text{ is true}) \leq 2 \exp\{\mu(s^*)\}.$$

b) For any rule for deciding between H' and H'', on the basis of the measurements $x_1, ..., x_N$, we have

$$P(\text{decide } H' \mid H'' \text{ is true}) + P(\text{decide } H'' \mid H' \text{ is true}) \geq \frac{1}{2} \exp\{\mu(s^*) - [2\mu''(s^*)]^{1/2}\},$$

where a prime indicates differentiation with respect to s.

Proof: Part (a) of the Lemma is the Corollary in p.84 of [SGB]. For part (b), it is shown in [SGB] (equation (3.42), p.87) that

$$P(\text{decide } H' \mid H'' \text{ is true}) + P(\text{decide } H'' \mid H' \text{ is true}) >$$

$$\frac{1}{2}\exp\{\mu(s) - s\mu'(s) - s[2\mu''(s)]^{1/2}\} + \frac{1}{2}\exp\{\mu(s) + (1-s)\mu'(s) - (1-s)[2\mu''(s)]^{1/2}\}, \quad \forall s \in (0,1).$$

If $s^* \in (0,1)$, we have $\mu'(s^*) = 0$ and the desired result follows immediately. If $s^* = 0$, we may take the limit in the above inequality, as $s \downarrow 0$. Since μ'' is continuous, and therefore bounded, we have $\lim_{s \downarrow 0} s\mu''(s) = 0$, which yields

$$P(\text{decide } H' \mid H'' \text{ is true}) + P(\text{decide } H'' \mid H' \text{ is true}) \geq \frac{1}{2}\exp\{\mu(0)\} \geq \exp\{\mu(0) - [2\mu''(0^*)]^{1/2}.$$

The last inequality follows because μ is convex and therefore $\mu''(s) \geq 0$, $\forall s$. The argument for the case $s^* = 1$ is identical. •

The bounds of parts (a) and (b) of the Lemma could be far apart if μ'' is left uncontrolled. For this reason we introduce the following assumption:

Assumption 1: a) $|\mu_{ij}(\gamma, s)| < \infty$, $\forall \gamma \in \Gamma$, $\forall i \neq j$, $\forall s \in [0,1]$.

b) There exists a constant A such that $|\mu''_{ij}(\gamma, s)| \leq A$, $\forall s \in [0,1]$, $\forall \gamma \in \Gamma$, $\forall i \neq j$.

The content of this Assumption is explored in Section VI; it is shown there that it corresponds to some minor restrictions on the distribution of the observations, which are satisfied in typical situations of practical interest.

As a preview of the remainder of the proof, we use Lemma 1, for each pair of distinct hypotheses to argue that the decision rules $\gamma_1, \ldots, \gamma_N$ of the sensors should be chosen so as to minimize

$$\max_{\{(i,j):\, i \neq j\}} \min_{s \in [0,1]} \sum_{k=1}^{N} \mu_{ij}(\gamma_k, s).$$

We reformulate this as a linear programming problem and use linear programming theory to show that a small number of different γ_k's suffices.

Let $\mathcal{F}$ be the set of all finite subsets of Γ. For any $F \in \mathcal{F}$, let

$$\Lambda(F) = \min_{x_\gamma} \max_{\{(i,j):\, i \neq j\}} \min_{s \in [0,1]} \sum_{\gamma \in F} x_\gamma \mu_{ij}(\gamma, s),$$

where the minimization with respect to x_γ is subject to the constraints

$$x_\gamma \geq 0, \qquad \forall \gamma \in F, \tag{2a}$$

$$\sum_{\gamma \in F} x_\gamma = 1. \tag{2b}$$

Let

$$\Lambda^* = \inf_{F \in \mathcal{F}} \Lambda(F).$$

Let us fix N and some collection $\gamma^N \in \Gamma^N$ of decision rules. Let $\alpha = \min_i P(H_i)$. We then have, using part (b) of Lemma 1,

$$J_N(\gamma^N) = \sum_{\{(i,j):\, i \neq j\}} P(\text{decide } \mathrm{H_i} \,|\, \mathrm{H_j}) \mathrm{P}(\mathrm{H_j}) \geq$$

$$\frac{\alpha}{2} \max_{\{(i,j):\, i \neq j\}} \exp \left[\sum_{k=1}^{N} \mu_{ij}(\gamma_k, s_{ij}^*) - \left(2 \sum_{k=1}^{N} \mu_{ij}''(\gamma_k, s_{ij}^*) \right)^{1/2} \right],$$

where s_{ij}^* minimizes $\sum_{k=1}^{N} \mu_{ij}(\gamma_k, s)$ over $s \in [0,1]$. Let F be the set of different decision rules (elements of Γ) which are present in the collection γ^N of decision rules. For each $\gamma \in F$, let x_γ be the proportion of the sensors using decision rule γ; that is x_γ is equal to the number of k's such that $\gamma_k = \gamma$, divided by N. By construction, the coefficients x_γ satisfy the constraints (2a-2b). Using Assumption 1b to bound $\mu_{ij}''(\gamma_k, s_{ij}^*)$, the definition of s_{ij}^* and the definition of $\Lambda(F)$, we have

$$J_N(\gamma^N) \geq \frac{\alpha}{2} \exp \left(\max_{\{(i,j):\, i \neq j\}} \min_{s \in [0,1]} \left[N \sum_{\gamma \in F} x_\gamma \mu_{ij}(\gamma, s) \right] - (2NA)^{1/2} \right) \geq$$

$$\frac{\alpha}{2} e^{N\Lambda(F) - (2NA)^{1/2}} \geq \frac{\alpha}{2} e^{N\Lambda^* - (2NA)^{1/2}}.$$

This shows that $R_N \geq \Lambda^* - (2A/N)^{1/2} + \frac{1}{N} \log(\alpha/2)$. Taking the limit as $N \to \infty$, we obtain

$$\liminf_{N \to \infty} R_N \geq \Lambda^*. \tag{3}$$

Lemma 2: $\Lambda^* = \inf_{F \in \mathcal{F}_0} \Lambda(F)$, where $\mathcal{F}_0$ is the collection of all subsets of Γ of cardinality no larger than $M(M-1)/2$.

Proof: Given some $F \in \mathcal{F}$, let s^*_{ij}, x^*_γ be such that the constraints (2a), (2b) are satisfied and

$$\Lambda(F) = \max_{\{(i,j):\ i \neq j\}} \sum_{\gamma \in F} x^*_\gamma \mu_{ij}(\gamma, s^*_{ij}).$$

(Such s^*_{ij}, x^*_γ exist because the quantity $\max_{\{(i,j):\ i \neq j\}} \sum_{\gamma \in F} x_\gamma \mu_{ij}(\gamma, s_{ij})$ is continuous in s_{ij}, x_γ and is defined over a compact set; therefore, the minimum arising in the definition of $\Lambda(F)$ is attained.) In particular, if the s^*_{ij}'s are fixed, then the x^*_γ's are determined by minimizing $\max_{\{(i,j):\ i \neq j\}} \sum_{\gamma \in F} x_\gamma \mu_{ij}(\gamma, s^*_{ij})$, subject to the constraints (2a)–(2b). This minimization is equivalent to the following linear programming problem:

$$\min\ \lambda$$

subject to

$$\lambda \geq \sum_{\gamma \in F} x_\gamma \mu_{ij}(\gamma, s^*_{ij}), \quad \forall i, j,\ i \neq j,$$

$$x_\gamma \geq 0, \quad \forall \gamma \in F,$$

$$\sum_{\gamma \in F} x_\gamma = 1.$$

Let T be the cardinality of the set F. The above defined linear program has $T+1$ variables and $T+1+M(M-1)/2$ constraints. From linear programming theory [PS], we know that there exists an optimal solution at which the number of constraints for which equality holds, is no smaller than the number of variables. Therefore, with this optimal solution at most $M(M-1)/2$ of the constraints hold with a strict inequality, which implies that at most $M(M-1)/2$ of the x_γ's are nonzero. Therefore, for any $F \in \mathcal{F}$ there exists some $F' \in \mathcal{F}_0$ such that $\Lambda(F') \leq \Lambda(F)$ This completes the proof of Lemma 2. •

Let us fix some N and some $\epsilon > 0$. Let F be a subset of Γ of cardinality no larger than $M(M-1)/2$ (that is, $F \in \mathcal{F}_0$), such that $\Lambda(F) \leq \Lambda^* + \epsilon$, which exists, because of Lemma 2. Let x^*_γ, and s^*_{ij} be such that

$$\max_{\{(i,j):\ i \neq j\}} \sum_{\gamma \in F} x^*_\gamma \mu_{ij}(\gamma, s^*_{ij}) = \Lambda(F) \leq \Lambda^* + \epsilon.$$

We now define a collection γ^N of decision rules to be used by the N sensors: for each $\gamma \in F$, we let exactly $\lfloor Nx_\gamma^* \rfloor$ of them use the decision rule γ; if there are any remaining sensors, which is the case if Nx_γ^* is not an integer for some γ, we let these sensors use an arbitrary decision rule out of the set F. Let N_0 be the number of these remaining sensors.

We now estimate the probability of error under this particular γ^N. The probability of error is bounded above by the probability of error for the case where the fusion center chooses to ignore the messages transmitted by the last N_0 sensors and this is what we will assume. We now have

$$J_N(\gamma^N) \leq \sum_{\{(i,j):\ i \neq j\}} P(\text{decide } H_i \mid H_j \text{ is true}) P(H_j) \leq$$

$$M^2 \max_{\{(i,j):\ i \neq j\}} [P(\text{decide } H_i \mid H_j \text{ is true}) + P(\text{decide } H_j \mid H_i \text{ is true})]. \tag{4}$$

The expression inside the brackets in the right hand side of (4) refers to the probabilities of error for a context in which H_i and H_j are the only hypotheses. Since the fusion center uses the MAP rule, it is using a decision rule which would be optimal even if it had to discriminate only between the two hypotheses H_i and H_j (always assuming that the last N_0 messages are ignored). Thus, for each pair of hypotheses, the upper bound on the probability of error furnished by Lemma 1(a) is applicable. This yields

$$J_N(\gamma^N) \leq 2M^2 \max_{\{(i,j):\ i \neq j\}} \exp \left[\sum_{\gamma \in F} \lfloor Nx_\gamma^* \rfloor \mu_{ij}(\gamma, s_{ij}^*) \right]. \tag{5}$$

We now use the inequality $Nx_\gamma^* - \lfloor Nx_\gamma^* \rfloor \leq 1$ to obtain

$$\sum_{\gamma \in F} \lfloor Nx_\gamma^* \rfloor \mu_{ij}(\gamma, s_{ij}^*) \leq \sum_{\gamma \in F} Nx_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) + \sum_{\gamma \in F} |\mu_{ij}(\gamma, s_{ij}^*)| \leq \sum_{\gamma \in F} Nx_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) + K,$$

where K is a constant independent of N. We substitute the above inequality in the right hand side of (5), then take logarithms and divide by N to obtain

$$Q_N \leq \frac{\log J_N(\gamma^N)}{N} \leq \frac{2 \log M}{N} + \frac{\log K}{N} + \max_{\{(i,j):\ i \neq j\}} \sum_{\gamma \in F} x_\gamma^* \mu_{ij}(\gamma, s_{ij}^*) \leq \Lambda^* + \epsilon + \frac{K'}{N},$$

where K' is another constant independent of N. We take the limit as $N \to \infty$ and use the fact that ϵ was arbitrary to conclude that $\limsup_{N \to \infty} Q_N \leq \Lambda^*$. We combine this inequality with (3) and the obvious inequality $R_N \leq Q_N$ to complete the proof of the theorem. •

III. SPECIAL CASES AND COMPUTATIONAL CONSIDERATIONS.

Let us start by stressing that the proof of Theorem 1 is constructive and suggests a procedure for determining an asymptotically optimal set of decision rules. Namely, we have to solve the optimization problem defining Λ^*. The value of Λ^* is the optimal exponent and the associated optimal values of the x_γ's are the proportions of the sensors who should use each decision rule γ.

Theorem 1 is most useful in the case of binary hypotheses ($M = 2$) and binary messages ($D = 2$). For that case it is known [TS] that, without any loss of optimality, we may assume that each sensor decides what to transmit by performing a likelihood ratio test, with an appropriate threshold. We thus let Γ be the set of all such decision rules. Furthermore, in this case we have $M(M-1)/2 = 1$ and Theorem 1 implies that it is asymptotically optimal to let every sensor use the same threshold. In order to compute Λ^* we only need to optimize over all subsets of Γ of cardinality 1. Therefore, the optimal threshold may be computed by solving the optimization problem

$$\min_{\gamma \in \Gamma} \min_{s \in [0,1]} \mu_{12}(\gamma, s). \tag{6}$$

Notice that each $\gamma \in \Gamma$ can be described by a single real number, the value of the threshold being employed. We are therefore dealing with a nonlinear optimization problem in two dimensions. In typical problems, the probabilities $p_i^\gamma(d)$ are given by simple analytical expressions, as a function of the threshold corresponding to γ. Therefore, simple analytical expressions are also available for $\mu_{12}(\gamma, s)$ as well. It is known that $\mu_{12}(\gamma, s)$ is a convex function of s, for every γ [SGB], which makes the optimization with respect to s easier. Unfortunately, we are not aware of any simple but nontrivial examples in which the solution of the above optimization problem and the corresponding value of the optimal threshold may be obtained analytically.

In the case of binary hypotheses ($M = 2$) and messages of arbitrary cardinality $D > 2$, it is known that likelihood ratio tests are again optimal except that each decision rule consists of $D-1$ thresholds which determine which one of the D messages is to be sent. The same discussion as for the case of $D = 2$ applies here and (asymptotically) each sensor should use the same set of thresholds. The only difference is that γ is parametrized by a $(D-1)$–dimensional real vector (as

opposed to a scalar). Thus, the problem (6), which needs to be solved in order to determine the optimal thresholds, is a D–dimensional optimization problem. This may become quite hard unless D is small, the reason being that, in general, $\mu(\gamma, s)$ is not a convex function of the parameters specifying γ.

For the case where $M > 2$, Theorem 1 is less useful for computing an asymptotically optimal set of decision rules. The reason is that we have to perform an optimization problem over all subsets of Γ of cardinality $M(M-1)/2$. In principle, it seems possible to reformulate the optimization problem defining Λ^* in a way that avoids having to consider each such subset of Γ (which would be impossible anyway if Γ is infinite). Namely, we might perform the minimization

$$\min_{x \in P} \max_{\{(i,j):\, i \neq j\}} \min_{s \in [0,1]} \int_{\Gamma} \mu_{ij}(\gamma, s)\, dx(\gamma),$$

where $x(\cdot)$ is a positive measure on Γ with $x(\Gamma) = 1$ and where P is the set of all such measures. Leaving aside the technical difficulties in showing that this is an equivalent problem, it still does not seem particularly promising from a computational point of view. It appears that the only cases in which a numerical solution is possible are those cases in which the set Y is finite and has small cardinality, because in that case Γ is also finite and has small cardinality. Notice that if $F_1 \subset F_2$, then $\Lambda(F_2) \leq \Lambda(F_1)$. Therefore, if Γ is finite, we have $\Lambda^* = \Lambda(\Gamma)$. This suggests that in order to compute Λ^* it is preferable to ignore Theorem 1: instead of computing $\Lambda(F)$ for each F of cardinality $M(M-1)/2$, and then taking the minimum, we may just compute $\Lambda(\Gamma)$.

An Example: Let $M = 3$, $D = 2$ and let $Y = \{1, 2, 3\}$. Let each hypothesis be equally likely and let the statistics of the observation y be as follows: conditioned on H_i being true, y takes the value i with probability $1 - 2\epsilon$ and takes each one of the remaining two values with probability ϵ $(0 < \epsilon < 1/4)$. There are three possible decision rules. The i-th possible decision rule is: $\gamma_i(y) = 1$ if and only if $y = i$. Notice that γ_1 does not provide any information useful in discriminating between H_2 and H_3. Thus, $\mu_{23}(\gamma_1, s) = 0$, $\forall s$; similarly, $\mu_{12}(\gamma_3, s) = \mu_{13}(\gamma_2, s) = 0$, $\forall s$. Furthermore, by symmetry, $\mu_{12}(\gamma_1, s) = \mu_{13}(\gamma_1, s) = \mu_{23}(\gamma_2, s)$, etc. Let α be the value of the minimum of $\mu_{12}(\gamma_1, s)$, over $s \in [0, 1]$. Let x_i be the proportion of sensors using γ_i. The optimal values of

x_1, x_2, x_3 are determined by solving the problem

$$\alpha \max_{x_1,x_2,x_3} \{x_1 + x_2, x_1 + x_3, x_2 + x_3\},$$

over the unit simplex. It is easy to see that the optimal solution is $x_1 = x_2 = x_3 = \frac{1}{3}$, exactly as expected from the symmetry of the problem, and the corresponding value of the optimal exponent Λ^* is $2\alpha/3$.

IV. ALTERNATIVE INTERPRETATIONS.

Theorem 1 may be restated in a different language refering to a different context. For simplicity, we only consider the case $M = 2$. Suppose that we want to transmit a binary message and that we have a collection of noisy, memoryless and independent channels in our disposal. We are allowed to transmit a total of N times using any of the available channels each time. A receiver observes the N outputs of the channels, uses its knowledge of which channels were being used, and makes a decision on what was transmitted. The problem consists of finding which channels should be used and how many times each, in order to maximize the probability of correct decoding. For small N, it may be better to use a different channel each time, even if the original message is binary. However, our result states that, for binary messages, as $N \to \infty$, there is a single best channel which should be used for all transmissions. To see the analogy, think of the hypothesis H_1 or H_2 as the value of the binary message which we want to transmit and think of u_i as the output of the i-th transmission. A different channel corresponds to a different decision rule and the characteristics of the channel correspond to the quantities $p_i^j(d)$.

A different analogy may be made in the context of optimal design of measurements for failure detection. Suppose that we have a system which may be in one of two states: up or down. We have a collection of devices which may be used for failure detection. They are, however, unreliable and may make errors of both types. Furthermore, the probabilities of either type of error can be different for different devices. Suppose that, in order to increase reliability we want to use N such devices. Then, our result states that, as $N \to \infty$, there exists a single best device and that we should use N replicas of it, rather than using many devices with different characteristics.

V. THE CONTENT OF ASSUMPTION 1.

In this section we explore Assumption 1. Our objective here is to obtain conditions on the distributions P_i under which Assumption 1 can be shown to hold. Proposition 1 below deals with Assumption 1(a).

Proposition 1: Assumption 1(a) fails to hold if and only if there are two hypotheses H_i, H_j, such that the corresponding measures P_i and P_j are mutually singular.†

Proof: Suppose that Assumption 1(a) fails. Then, there exist some i, j and some $\gamma \in \Gamma$ for which $p_i^\gamma(d)p_j^\gamma(d) = 0$, $\forall d \in \{1,...,D\}$. Thus, for any $d \in \{1,...,D\}$, the set $\{y \in Y : \gamma(y) = d\}$ has non–zero measure under P_i only if it has zero measure under P_j. Since the sets $\{y \in Y : \gamma(y) = d\}$ cover the entire set Y, it follows that P_i and P_j are mutually singular. •

As a consequence of Proposition 1, we can see that if there are only two hypotheses and Assumption 1(a) fails to hold we are dealing with the uninteresting situation where each sensor is able to determine the true hypothesis on its own, with zero probability of error. For the case of more than two hypotheses, however, there are nontrivial detection problems in which Assumption 1a fails to hold. We conjecture that a somewhat modified version of Theorem 1, covering such a case, is possible. We now explore Assumption 1(b) and show that it holds for two interesting situations.

Proposition 2: Suppose that the observation set Y is finite and that Assumption 1(a) holds. Then Assumption 1(b) also holds.

Proof: The derivatives of $\mu_{ij}(\gamma, s)$, with respect to s are easily calculated to be [SGB, equations (3.24)–(3.25)]:

$$\mu'_{ij}(\gamma, s) = \sum_{d=1}^{D} \frac{(p_i^\gamma(d))^{1-s}(p_j^\gamma(d))^s}{\sum_{c=1}^{D}(p_i^\gamma(c))^{1-s}(p_j^\gamma(c))^s} \log \frac{p_j^\gamma(d)}{p_i^\gamma(d)}, \tag{7}$$

$$\mu''_{ij}(\gamma, s) = \left[\sum_{d=1}^{D} \frac{(p_i^\gamma(d))^{1-s}(p_j^\gamma(d))^s}{\sum_{c=1}^{D}(p_i^\gamma(c))^{1-s}(p_j^\gamma(c))^s} \left(\log \frac{p_j^\gamma(d)}{p_i^\gamma(d)}\right)^2\right] - [\mu'_{ij}(\gamma, s)]^2, \tag{8}$$

where all summations are made over those c's and d's for which $p_i^\gamma(c)p_j^\gamma(c)$, (respectively, $p_i^\gamma(d)p_j^\gamma(d)$) is nonzero.

† Two positive measures P_1, P_2, defined on a common (measurable) space Y are called mutually singular if there exists a measurable subset U of Y such that $P_1(U) = P_2(Y - U) = 0$.

Let α be the minimum of $p_i^\gamma(c)$, where the minimum is taken over all choices of γ, c, i, such that $p_i^\gamma(c) > 0$. Since Y is finite, the set of all possible decision rules γ is also finite and therefore α is the minimum of finitely many positive quantities and is itself positive. By Assumption 1(a) the denominator in equation (7) must have a nonzero summand and this summand will be bounded below by $\alpha^{1-s}\alpha^s = \alpha$. The numerator is bounded by D. Concerning the logarithmic term, it is bounded, in absolute value, by $|\log \alpha|$, for any d in the range of the summation. We conclude that $\mu'_{ij}(\gamma, s)$ is bounded in absolute value by a constant independent of i, j, γ, s. A similar argument applies to $\mu''(\gamma, s)$ and concludes the proof. •

Proposition 3: Suppose that, for any i, j, the measure P_i is absolutely continuous with respect to P_j and let L_{ij} denote the Radon–Nikodym derivative dP_i/dP_j. Assume that

$$E_i[\log^2 L_{ij}] < \infty, \qquad \forall i, j. \tag{9}$$

Then Assumption 1 holds.

Proof: The fact that Assumption 1(a) holds is immediate from our assumption of absolute continuity and Proposition 1.

For any decision rule $\gamma : Y \mapsto \{1, ..., D\}$, let $\mathcal{F}^\gamma$ be the smallest σ–field contained in $\mathcal{F}$ with respect to which the function γ is measurable. Let P_i^γ denote the restriction of the measure P_i on the σ-field $\mathcal{F}^\gamma$. It follows from the absolute continuity assumption that P_i^γ is absolutely continuous with respect to P_j^γ. We define L_{ij}^γ to be equal to the Radon–Nikodym derivative dP_i^γ/dP_j^γ. As is well known

$$L_{ij}^\gamma = E_j[L_{ij} \mid \mathcal{F}^\gamma], \quad \text{a.s. } (P_j). \tag{10}$$

Consider the function $\phi : (0, \infty) \mapsto (0, \infty)$ defined by $\phi(t) = t \log^2 t$. An easy calculation shows that it is convex. Therefore, using (10) and Jensen's inequality,

$$E_i[\log^2 L_{ij}^\gamma] = E_j[L_{ij}^\gamma \log^2 L_{ij}^\gamma] = E_j[\phi(L_{ij}^\gamma)] = E_j[\phi(E_j[L_{ij} \mid \mathcal{F}^\gamma])] \leq$$

$$E_j[E_j[\phi(L_{ij} \mid \mathcal{F}^\gamma]] = E_j[L_{ij} \log^2 L_{ij}] = E_i[\log^2 L_{ij}].$$

Using (9), we conclude that there exists a constant $B < \infty$ such that $E_i[\log^2 L_{ij}^\gamma] \leq B$, $\forall \gamma, i, j$; using the inequality $E[|x|] \leq 1 + E[x^2]$, we obtain the same conclusion for $E_i[\log L_{ij}^\gamma]$.

Notice now that $L^\gamma_{ij}(y) = p^\gamma_i(d)/p^\gamma_j(d)$, for every y such that $\gamma(y) = d$, almost surely. Using this observation, equation (7) may be rewritten as

$$\mu'_{ij}(\gamma, s) = \frac{E_i[(L^\gamma_{ji})^s \log L^\gamma_{ji}]}{E_i[(L^\gamma_{ji})^s]}; \tag{11}$$

similarly, equation (8) becomes

$$\mu''_{ij}(\gamma, s) = \frac{E_i[(L^\gamma_{ji})^s \log^2 L^\gamma_{ji}]}{E_i[(L^\gamma_{ji})^s]} - [\mu'_{ij}(\gamma, s)]^2. \tag{12}$$

Using the obvious inequality $(L^\gamma_{ji})^s \leq (1 + L^\gamma_{ji})$, $\forall s \in [0,1]$, we obtain the bound

$$|\mu'_{ij}(\gamma, s)| \leq \frac{|E_i[\log L^\gamma_{ji}]| + |E_i[L^\gamma_{ji} \log L^\gamma_{ij}]|}{E[(L^\gamma_{ji})^s]} = \frac{|E_i[\log L^\gamma_{ji}]| + |E_j[\log L^\gamma_{ij}]|}{E[(L^\gamma_{ji})^s]}.$$

We have already proved that the numerator is bounded. We now establish a lower bound on $E_i[(L^\gamma_{ji})^s]$. Since $E_i[L_{ji}] = 1$, it follows that there exists a $\mathcal{F}$–measurable set $Y_0 \subset Y$ and some $\epsilon > 0$, $\delta > 0$, such that $P_i(Y_0) \geq \epsilon$ and $L_{ji}(y) \geq \delta$, $\forall y \in Y_0$. Since $x^s \geq \min\{1, x\}$, we obtain $E_i[L^s_{ji}] \geq \epsilon \min\{1, \delta\}$, $\forall s \in [0,1]$. We now use the fact that the function $\phi(x) = x^s$ is concave, for any fixed $s \in [0,1]$, and Jensen's inequality to obtain

$$E_i[(L^\gamma_{ji})^s] = E_i[(E_i[L_{ji} \mid \mathcal{F}^\gamma])^s] \geq E_i[E_i[L^s_{ji} \mid \mathcal{F}^\gamma]] = E_i[L^s_{ji}] \geq \epsilon \min\{1, \delta\}.$$

This concludes the proof that $\mu'(\gamma, s)$ is bounded. The proof of the boundedness of $\mu''(\gamma, s)$ is identical and is omitted. •

VI. THE NEYMAN–PEARSON PROBLEM.

In this section we consider the Neyman–Pearson version of the problem studied in the preceding sections. We are given an observation set Y, endowed with a σ-field $\mathcal{F}$. There are two hypotheses ($M = 2$) and for each hypothesis we are given a measure P_i on $(Y, \mathcal{F})$, $i = 1, 2$. Let D be a fixed positive integer and let Γ be the set of all measurable functions $\gamma : Y \mapsto \{1, ..., D\}$. As before, the i-th sensor makes an independent observation y_i whose statistics are described by P_j, assuming that hypothesis H_j is true. Again, the i-th sensor transmits a message $\gamma_i(y_i)$ to a fusion center, where $\gamma_i \in \Gamma$, and finally the fusion center makes a final decision using a decision rule γ_0. We allow

γ_0 to be randomized. That is, the final decision of the fusion center may depend on the messages it has received as well as an internally generated random variable. Let Γ_0 be the set of all candidate decision rules γ_0 for the fusion center.

For any given $(\gamma_0, \gamma_1, ..., \gamma_N) \in \Gamma_0 \times \Gamma^N$, consider the probabilities of error defined by

$$J_N^1(\gamma_0, \gamma_1, ..., \gamma_N) = P_1(\gamma_0(\gamma(y_1), \cdots \gamma(y_N)) = 2), \tag{13}$$

$$J_N^2(\gamma_0, \gamma_1, ..., \gamma_N) = P_2(\gamma_0(\gamma(y_1), \cdots \gamma(y_N)) = 1). \tag{14}$$

Let us fix a constant β belonging to $(0,1)$. We would like to minimize $J_N^1(\gamma_0, ..., \gamma_N)$, over all $\gamma_0, ..., \gamma_N$ satisfying

$$J_N^2(\gamma_0, \gamma_1, \cdots, \gamma_N) \leq 1 - \beta. \tag{15}$$

The optimal value of J_N^1 falls exponentially with N and we define

$$r_N(\gamma_0, \cdots, \gamma_N) = \frac{1}{N} \log J_N^1(\gamma_0 \cdots \gamma_N).$$

Let

$$R_N = \inf r_N(\gamma_0, \cdots, \gamma_N), \tag{16}$$

where the infimum is taken over all $(\gamma_0, \cdots, \gamma_N) \in \Gamma_0 \times \Gamma^N$ satisfying (15). We will use the following assumption:

Assumption 2: a) P_2 is absolutely continuous with respect to P_1;

b)

$$E_2 \left[\log^2 \left(\frac{dP_2}{dP_1} \right) \right] = A < \infty, \tag{17}$$

where dP_2/dP_1 is the Radon–Nikodym derivative of the two measures.

We define $\mathcal{F}^\gamma$ and P_i^γ as in Section V: $\mathcal{F}^\gamma$ is the σ–field on Y generated by γ and P_i^γ is the measure P_i restricted to $\mathcal{F}^\gamma$. The argument in the proof of Proposition 3, in Section V, applies here and shows that $E_2[\log^2(dP_2^\gamma/dP_1^\gamma)] \leq A$, $\forall \gamma \in \Gamma$. The latter inequality also implies that there exists some $B < \infty$ such that

$$K(\gamma) = E_2 \left[\log \frac{dP_2^\gamma}{dP_1^\gamma} \right] \leq B, \qquad \forall \gamma \in \Gamma. \tag{17}$$

The quantity $K(\gamma)$ defined by equation (18) may be recognized as the Kullback–Liebler [KL] information distance between the distributions of the random variable $\gamma(y)$ under the two alternative hypotheses. It is guaranteed to be nonnegative. Furthermore, Stein's Lemma [B] states that $K(\gamma)$ is the asymptotic error exponent if all sensors are using the same decision rule γ and if the fusion center chooses γ_0, according to the Neyman–Pearson Lemma. In light of this fact, the following result should be expected.

Theorem 2: If Assumption 2 holds, then

(i) $\lim_{N\to\infty} R_N = -\sup_{\gamma\in\Gamma} K(\gamma)$.

(ii) The value of R_N stays the same if in the definition (16) we impose the additional constraint $\gamma_1 = \cdots = \gamma_N$.

Proof: (Outline) Fix some $\epsilon > 0$ and let $\gamma^* \in \Gamma$ be such that $K(\gamma^*) \geq \sup_{\gamma\in\Gamma} K(\gamma) - \epsilon$. Let the fusion center choose γ_0 optimally, subject to (15). From Stein's Lemma, we obtain $\lim_{N\to\infty} r_N(\gamma_0, \gamma^*, \cdots, \gamma^*) = -K(\gamma^*)$. In particular, $\limsup_{N\to\infty} R_N \leq -K(\gamma^*) \leq -\sup_{\gamma\in\Gamma} K(\gamma) + \epsilon$. Since ϵ was arbitrary, we conclude that $\limsup_{N\to\infty} R_N \leq -\sup_{\gamma\in\Gamma} K(\gamma)$ and we have shown this bound to be valid under the additional constraint $\gamma_1 = \cdots \gamma_N$.

In order to complete the proof, it is sufficient to show that for any $\gamma_0, ..., \gamma_N$ satisfying (15) we have

$$r_N(\gamma_0, \cdots, \gamma_N) \geq -\frac{1}{N}\sum_{i=1}^{N} K(\gamma_i) + f(N) \geq -\sup_{\gamma\in\Gamma} K(\gamma) + f(N), \tag{19}$$

where f is a function with the property $\lim_{N\to\infty} f(N) = 0$ and which does not depend on $\gamma_0, \cdots, \gamma_N$. While this result does not follow from the usual formulation of Stein's Lemma (which uses the Assumption $\gamma_1 = \cdots = \gamma_N$), it may be proved by a small variation of the proof of that Lemma, and for this reason the proof is omitted. Suffice to say that we may take the proof of Stein's Lemma given in [B]. Wherever in that a proof convergence in probability of a log–likelihood ratio to its mean is asserted, we replace such a statement with an inequality which bounds the probability of a deviation of a log–likelihood ratio from its mean. Such an inequality is obtained from Chebychev's inequality. Because of (17) the variance of the log–likelihoods of interest admits the same bound, irrespective of the choice of the γ_i's. For this reason, the function f in (19) may

be taken independent of the γ's. The proof is then completed by taking the infimum of both sides of (19), over all $\gamma_0, ..., \gamma_N$ and then letting N tend to infinity. •

We continue with a few observations. For simplicity we restrict our discussion to the case of binary messages ($D = 2$).

It is easy to prove that there is no loss of optimality if we constrain the γ_i's to correspond to likelihood ratio tests [HV]. If we are only interested in asymptotics, the same conclusion may be obtained from Theorem 2: it is not hard to show that if a decision rule does not have the form of a likelihood ratio test, then another decision rule can be found for which $K(\gamma)$ is even larger. This leads to the conclusion that asymptotically optimality is not lost by assuming that each γ_i consists of a comparison of the likelihood ratio computed by that sensor with a threshold.

As is well–known, randomization is generally required in optimal hypothesis testing, under the Neyman–Pearson formulation. For this reason, we allowed the decision rule of the fusion center to employ an internally generated random variable. We may ask whether anything can be gained by allowing the sensors as well to use randomized decision rules. The answer is generally positive. For example, if $N = 1$, then the best strategy is to let the single sensor perform an optimal Neyman–Pearson test (for which randomization is needed) and have the fusion center adopt the decision of the sensor. Interestingly enough, however, randomization does not help asymptotically as $N \to \infty$, which we now prove. For any two measures P, Q on $(Y, \mathcal{F})$, let $K(Q, P) = E[\log(dQ/dP)]$, where the expectation is with respect to Q. With this notation, $K(\gamma) = K(P_2^\gamma, P_1^\gamma)$, $\forall \gamma \in \Gamma$. It is known, and easy to show, that $K(Q, P)$ is a convex function of (Q, P). Suppose now that a sensor uses a decision rule which involves randomization. The pair (P_2^γ, P_1^γ) of the probability distributions of the message transmitted by a sensor using a randomized decision rule γ lies in the convex hull of such pairs of probability distributions corresponding to non–randomized decision rules. Using the convexity pf K, it follows that randomization cannot help in increasing the supremum of $K(\gamma)$ and, therefore, does not help asymptotically.

From a computational point of view, the problem of this section is a little easier from the problem of Section II, the reason being that we do not have the additional free parameter s of Section II.

In particular, with decision rules parametrized by a scalar threshold, maximization of $K(\gamma)$ is equivalent to a one–dimensional optimization problem. As there may be multiple local optima, some form of exhaustive search may be required.

As an illustration, we study the performance of a naive selection of the decision rule γ of each sensor. We let each sensor perform a maximum likelihood test and transmit its decision to the fusion center. This is certainly a bad idea if $N = 1$ because in that the case the sensor should perform a Neyman–Pearson test which is, generally, different from a maximum likelihood test. Still, one may wonder whether such a naive prescription has any performance guarantees, as $N \to \infty$. The answer is negative, as the following example shows. Let P_1 and P_2 be as in Figure 1. A decision rule γ corresponding to a maximum likelihood test is to let $\gamma(y) = 1$ if and only if $y > 1/2$. For this choice of γ, if we assume that ϵ is small enough and use a Taylor series expansion we obtain

$$K(\gamma) = \frac{1}{2}\log\left(\frac{1/2}{\frac{1}{2}-\epsilon}\right) + \frac{1}{2}\log\left(\frac{1/2}{\frac{1}{2}+\epsilon}\right) \leq A\epsilon^2,$$

where A is some positive constant. Let us now consider the decision rule γ given by $\gamma(y) = 1$ if and only if $y > 1$. We then have $K(\gamma) = \log(1/(1-\epsilon/2)) \geq \epsilon/2 + B\epsilon^2$, for some constant B. We conclude from this example that the naive decision rule suggested above can be far from optimal (in terms of error exponent) by an arbitrary multiplicative factor.

ACKNOWLEDGEMENTS: The author is grateful to Professor Robert Gallager who suggested that the results of [SGB] could be used in proving Theorem 1.

REFERENCES

[B] Bahadur, R.R., *Some Limit Theorems in Statistics,* SIAM, Philadelphia, 1971.

[E] Ekchian, L.T., "Distributed detection and communication problems", Ph.D. dissertation, Dept. of EECS, MIT, Cambridge, MA, 1982.

[HV] Hoballah, I.Y., and Varshney, P.K., "Decentralized Neyman–Pearson Detection," *Proceedings of the 25th IEEE Conference on Decision and Control,* Athens, Greece, December 1986.

[KL] Kullback, S., and Leibler, R.A., "On Information and Sufficiency", *Annals of Mathematical*

Statistics, 22, 1951, pp. 279–280.

[LS] Lauer, G.S., and Sandell, N.R., Jr., "Decentralized detection given waveform observations", Alphatech Inc. Technical Paper TP–131, Burlington, MA, 1982.

[SGB] Shannon, C.E., Gallager, R.G., Berlekamp, E.R., "Lower bounds to error probability for coding on discrete memoryless channels, I", *Information and Control,* 10, 1967, pp. 65–103.

[PS] Papadimitriou, C.H.,. Steiglitz, K., *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall, 1982.

[R] Rudin, W., *Real and Complex Analysis,* McGraw Hill, 1974.

[TA] Tsitsiklis, J.N., Athans, M., "On the Complexity of Decentralized Decision Making and Detection Problems", *IEEE Transactions on Automatic Control,* AC–30, 5, 1985.

[TS] Tenney, R.R., Sandell, N.R., Jr., "Detection with distributed sensors",*IEEE Transactions on Aerospace and Electronic Systems,* AES–17, 4, 1981.

APPENDIX

We consider here the problem introduced in Section II, with two hypotheses ($M = 2$), binary messages ($D = 2$), two sensors ($N = 2$), and with y_1, y_2 identically distributed and conditionally independent given either hypothesis. We present an example which shows that it is possible that different sensors may have to use different decision rules even if their observations are identically distributed. An example of this type was presented in [TeSa]. However, that example used a special cost function which introduced a large penalty if both sensors send the same message and the wrong decision is made by the fusion center. Naturally, this creates an incentive for the sensors to try to transmit different messages, and therefore use different decision rules. Thus, the asymmetry of the optimal decision rules of the two sensors can be ascribed to this particular aspect of the cost function and does not prove that asymmetrical decision rules may be optimal for our cost function (probability of error).

Our example is the following. We let H_1 and H_2 be equally likely. The observations y_1, y_2 are

conditionally independent, given either hypothesis, take values in $\{1,2,3\}$ and have the following common distribution:

$P(y=1|H_1)=4/5, \quad P(y=2|H_1)=1/5 \quad P(y=3|H_1)=0,$

$P(y=1|H_2)=1/3, \quad P(y=2|H_2)=1/3 \quad P(y=3|H_2)=1/3.$

An optimal set of decision rules may be found by exhaustive enumeration. Since each sensor has to perform a likelihood ratio test, there are only two candidate decision rules for each sensor:

(A) $u_i=1$ iff $y_i=1$,

(B) $u_i=1$ iff $y_i \in \{1,2\}$.

Thus, we need to consider three possibilities: (i) both sensors use (A); (ii) both sensors use (B); sensor 1 uses (A) and sensor 2 uses (B). Naturally, we assume that the fusion center is using the maximum a posteriori probability rule.

Explicit evaluation of the expected cost for each possibility shows that the optimal set of decision rules consists of one sensor using decision rule A, one sensor using decision rule B and the fusion center deciding H_1 if and only if $u_1=u_2=1$, for an expected cost of 19/90.

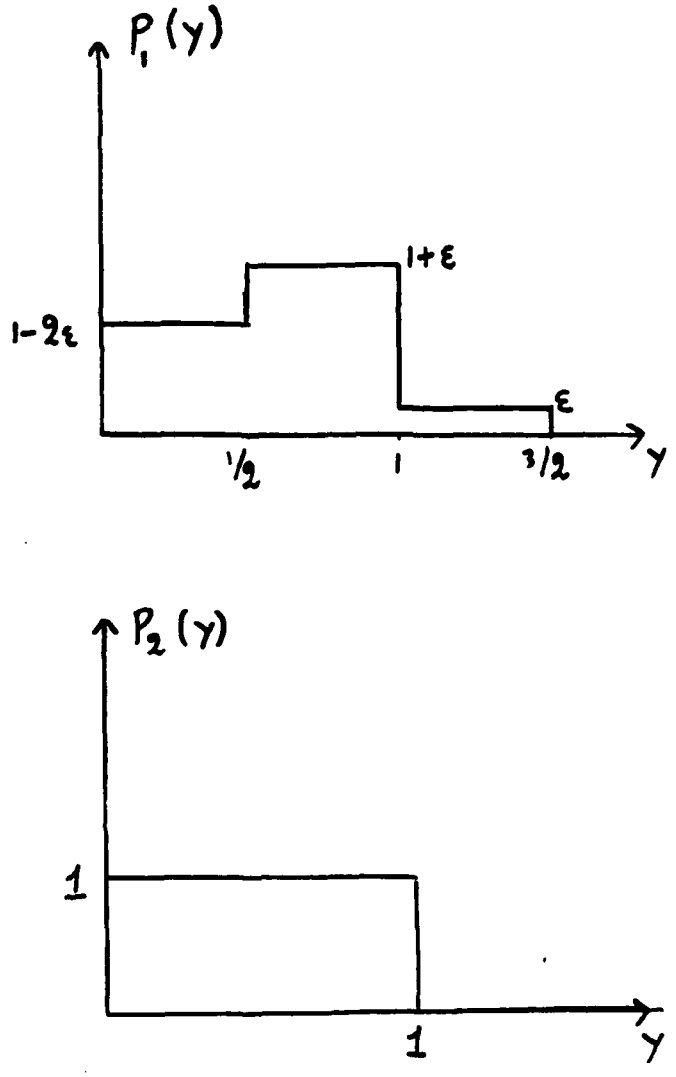


Fig. 1

END

7-87

DTIC